FirstDiscovery 3.0

User Manual

Copyright © 2004 Schrödinger, LLC. All rights reserved.
Schrödinger, FirstDiscovery, Glide, Impact, Jaguar, Liaison, LigPrep, Maestro, Prime, QSite, and QikProp are trademarks of Schrödinger, LLC.
MacroModel is a registered trademark of Schrödinger, LLC.
To the maximum extent permitted by applicable law, this publication is provided "as is" without warranty of any kind. This publication may contain trademarks of other companies.

Revision A, June 2004

Contents

Chapter 1: Introduction 1
1.1 Document Conventions 2
1.2 Introduction to FirstDiscovery Panels 2
1.2.1 Job Options 2
1.2.2 Panel Controls and Tabbed Folders 4
1.2.3 Start, Write, Hide, and Help 4

Chapter 2: Introduction to Maestro 5
2.1 General Interface Behavior 5
2.2 Starting Maestro 5
2.3 The Maestro Main Window 6
2.3.1 The Menu Bar 7
2.3.2 The Toolbar 8
2.3.3 Mouse Functions in the Workspace 11
2.3.4 Shortcut Key Combinations 11
2.4 Maestro Projects 12
2.4.1 The Project Table Toolbar 13
2.4.2 The Project Table Menus 15
2.4.3 Selecting Entries 15
2.4.4 Including Entries in the Workspace 16
2.4.5 Mouse Functions in the Project Table 16
2.4.6 Project Table Shortcut Keys 17
2.5 Building a Structure 18
2.6 Selecting Atoms 21
2.7 Maestro Command Scripts 22
2.8 Specifying a Maestro Working Directory 23
2.9 Undoing an Operation 24
2.10 Running and Monitoring Jobs 24
2.11 Getting Help 26
2.12 Ending a Maestro Session 26

Chapter 3: FirstDiscovery from the Command Line 27
3.1 Command-Line Capabilities 27
3.1.1 Location of Files and Working Directory 27
3.1.2 FirstDiscovery Job Types 27

3.1.3 Reasons to Run Jobs from the Command Line 27
3.1.4 Force Fields and Write Template 28
3.2 File Name Conventions 28
3.3 The Impact Command: Usage Summary 29
3.4 Running Jobs From the Command Line 31
3.4.1 Protein Preparation 32
3.4.2 Glide 32
3.4.3 Liaison 33
3.4.4 QSite 33
3.4.5 Basic Impact 34
3.4.6 Using Job Control Commands 34
3.5 Using Command-Line Utilities 35

Chapter 4: Protein Preparation 37

4.1 Protein and Ligand Structure Preparation 37
4.2 The ProteinPrep Panel 37
4.3 Step-by-Step Overview 38
4.4 Importing the Protein Complex Structure 39
4.5 Deleting Unwanted Waters 39
4.5.1 Locating Structural Waters 40
4.5.2 Should Structural Waters Be Kept? 41
4.5.3 Deleting All Water Molecules 41
4.5.4 Deleting Distant Water Molecules 41
4.5.5 Delete Remaining Unwanted Waters 42
4.6 Simplifying the Protein Complex 42
4.6.1 Determining Whether the Complex Is a Multimer 42
4.6.2 Retaining Needed Subunits 43
4.7 Adjusting the Protein, Metal Ions, and Cofactors 45
4.7.1 Proteins That Already Include Hydrogen Atoms 45
4.7.2 Checking the Protein Structure for Metal Ions and Cofactors 45
4.7.3 Adjusting Metal Ions 45
4.7.4 Displaying the Cofactor 47
4.7.5 Adjusting Cofactor Atom and Bond Properties 47
4.8 Adjusting the Ligand 49
4.8.1 Adjusting Ligand Atom and Bond Properties 49
4.8.2 Manually Deleting Explicit Ligand-Metal Bonds 50
4.8.3 Checking for Other Protein-Ligand Bonds 51

4.9 Running ProteinPrep on the Structures 51
4.9.1 Entering Job Settings 52
4.9.2 Defining the Ligand 52
4.9.3 Choosing a Procedure 52
4.9.4 Other Options 53
4.9.5 Launching the ProteinPrep Job 53
4.9.6 Output Job Files 53
4.10 Checking the Output Structures 54
4.10.1 Checking the Orientation of Water Molecules 54
4.10.2 Resolving H-Bonding Conflicts 55
4.11 Command-Line Protein Preparation 55
4.11.1 Usage Summary for `protprep` 55
4.11.2 Usage Summary for `pprep` 57
4.11.3 Usage Summary for `impref` 58

Chapter 5: Ligand Preparation 59
5.1 Ligand Preparation Checklist 59
5.2 LigPrep 59
5.2.1 The LigPrep Process 60
5.2.2 The LigPrep Panel 62
5.3 The Ionization State Expander (`ionizer`) 63

Chapter 6: Glide 65
6.1 Introduction to Glide 65
6.1.1 Glide Constraints 68
6.1.2 Glide Extra-Precision Mode 69
6.1.3 Glide/Prime Induced Fit 69
6.2 The Glide Panel 70
6.3 The Settings Folder 70
6.3.1 Glide Function: Set Up Grids or Dock Ligands 70
6.3.2 Docking Mode Options and Using Extra-Precision Mode 71
6.3.3 Other Settings Folder Options 72
6.4 The Site Folder 74
6.4.1 Site Folder Features 75
6.4.2 The Ligand to Define Grid Panel 78
6.4.3 The Active Site Residues Panel 79
6.5 The Ligand Folder 80
6.5.1 Ligand Folder Features 80
6.5.2 The Reference Ligand Panel 85

6.6 The Scoring Folder 86
6.6.1 Refinement of Initial Poses Section 88
6.6.2 Energy Minimization Section 89
6.6.3 Final Scoring Section 89
6.7 The Output Folder 90
6.7.1 Elimination of Duplicate Poses Section 90
6.7.2 Structure Output Section 91
6.7.3 Raw Pose Files, Output Pose Files, and `glide_sort` 92
6.8 The Constraints Folder 94
6.8.1 Using Glide Constraints 94
6.8.2 Glide Constraints Folder Features: Grid Generation 95
6.8.3 Glide Constraints Folder Features: Docking 97
6.9 The Similarity Folder 98
6.9.1 Introduction to Similarity Scoring in Glide 98
6.9.2 Similarity Folder Features: Grid Generation 99
6.9.3 Similarity Folder Features: Docking 101
6.10 The Pose Viewer 103
6.10.1 Pose Viewer Panel Features 103
6.10.2 The PoseWrite Panel 107
6.11 Glide Utilities 107
6.11.1 `glide_sort` 108
6.11.2 `glide_rescore` 109
6.11.3 `para_glide` 110

Chapter 7: Liaison 113

7.1 Brief Description of Liaison 113
7.2 Liaison Simulations 114
7.2.1 Liaison Directory Structure 114
7.2.2 Directories Created 114
7.2.3 Files Created 116
7.2.4 Liaison Simulation Requirements 117
7.3 Liaison Analysis 118
7.3.1 Fitting the Simulation Results to Experimental Data 118
7.3.2 Predicting Binding Affinities of New Ligands 119
7.4 The Liaison Panel 119
7.5 The Settings Folder 120
7.6 The System Folder 122
7.6.1 Multiple Ligands, Single Receptor 123
7.6.2 Single Ligand, Single Receptor 124

7.7 The Parameters Folder 125
7.7.1 Sampling Method 125
7.7.2 Ligand Simulation and Ligand/Receptor Simulation 127
7.8 The Constraints Folder 130
7.9 The Analysis Folder 132
7.9.1 Analysis Settings Section 132
7.9.2 Ligand Specification Section 134
7.10 Running Liaison as a Stand-Alone Program 136
7.11 Killing Liaison Jobs 137

Chapter 8: QSite 139
8.1 Using QSite 139
8.2 The QSite Panel 140
8.3 The Potential Folder 140
8.4 The Constraints Folder 142
8.5 The Minimization Folder 143
8.6 The Optimization Folder 144
8.7 The QM Settings Folder 147
8.7.1 QM Settings Folder Features 147
8.7.2 The QM Residues/Ligands Panel 149
8.7.3 The QM Region Ions Panel 151

Chapter 9: Energy Minimization 153
9.1 Basic Impact Applications 153
9.2 Using the Energy Minimization Panel 153
9.3 Energy Minimization Panel Features 154
9.4 The Potential Folder 154
9.4.1 Potential Folder Options 154
9.4.2 Potential Folder Methods 156
9.4.3 The Truncation Panel 157
9.4.4 The Fast Multipole Method Panel 157
9.4.5 The Periodic Boundary Conditions Panel 158
9.4.6 The Continuum Solvation Panel 159
9.5 The Constraints Folder 160
9.5.1 Constraints Folder Features 160
9.5.2 The Frozen Atoms Panel 161
9.5.3 The Buffered Atoms Panel 162
9.6 The Minimization Folder 162

Chapter 10: Molecular Dynamics Simulations 165
10.1 Using the Dynamics Panel 165
10.2 Dynamics Panel Features 165
10.3 The Constraints Folder 166
10.4 The MD Parameters Folder 167
10.5 The Dynamics Folder 170

Chapter 11: Hybrid Monte Carlo Simulations 173
11.1 Using the Hybrid Monte Carlo Panel 173
11.2 Impact Hybrid Monte Carlo Panel Features 173
11.3 The Hybrid MC Folder 174

Chapter 12: Soak—Add Explicit Water Solvent 177
12.1 Using the Soak Panel 177
12.2 Soak Panel Features 178

Chapter 13: Getting Help.....179

Index.....181

Chapter 1: Introduction

Once you have installed FirstDiscovery according to the instructions in the *Schrödinger Product Installation Guide,* the *FirstDiscovery User Manual* will help you use the FirstDiscovery software suite. Some of the material in this manual provides context and background information, but much of it describes the options and settings needed to configure FirstDiscovery and Basic Impact applications.

The FirstDiscovery suite includes Glide, Liaison, and QSite (the principal program modules), Basic Impact applications, Protein Preparation, and supporting utilities and scripts. The FirstDiscovery programs are run primarily from the Maestro graphical user interface, an introduction to which appears in Chapter 2, but can also be run from the command line, as described in Chapter 3. Utilities and scripts are run from the command line. Both interfaces call on Impact, the FirstDiscovery calculation engine.

For related information, please see our other documentation:

- The *Schrödinger Product Installation Guide*, which includes installation instructions for all Schrödinger products and documentation.
- The *FirstDiscovery Quick Start Guide*, which contains tutorials intended to familiarize you with protein preparation, Glide, Liaison, and QSite.
- The *FirstDiscovery Technical Notes*, featuring in-depth scientific discussions of Glide, Liaison, and QSite, including performance and usage notes.
- The *FirstDiscovery Command Reference Manual*, which contains syntax and keywords for Impact command input files.

  Starting with FirstDiscovery 3.0, the *Command Reference Manual* is no longer printed as hardcopy and bound with the other FirstDiscovery manuals. Like other FirstDiscovery documentation, it is still provided in PDF format, compressed into a tar file, on the FirstDiscovery CD, and is available at Schrödinger's support web page, http://www.schrodinger.com/Support/pdf.html
- The *Maestro User Manual*, which describes how to use the features of Maestro, including the Atom Selection Dialog (ASD). An appendix describes command-line utilities, many of which are used in FirstDiscovery.
- The *Maestro Command Reference Manual*, which contains commands, options, and arguments for running Maestro from the command line, including the Atom Specification Language (ASL) and the Entry Specification Language (ESL).

1.1 Document Conventions

In addition to the normal use of italics for names of documents, the font conventions that are used in this manual are summarized in Table 1.1.

Table 1.1. Font Conventions.

Font	Example	Use
Sans serif	Project Table	Names of GUI features such as panels, menus, menu items, buttons, labels
Monospace	`$SCHRODINGER/maestro`	File names, directory names, commands, and environment variables
Italics	*filename*	Text that the user must replace with a value
Sans serif uppercase	ALT+H	Keyboard keys

In descriptions of command syntax, the usual UNIX conventions are used: square brackets `[ ]` enclose optional items, braces `{ }` enclose required items, and the pipe symbol `|` separates items in a list from which one item must be chosen.

In this document, to *type* a command means to type the required text in the specified location, and to *enter* a command means to type the required text then press the RETURN key.

1.2 Introduction to FirstDiscovery Panels

FirstDiscovery panels in Maestro have a common structure, with standard features and options in the upper and lower regions of the panel.

1.2.1 Job Options

At the top of the panel are job options. The first three features are common to FirstDiscovery (Protein Preparation, Glide, Liaison, and QSite) as well as Impact panels.

Job

The default file name for jobs started from a panel is displayed in the Job text box, e.g. `glidetmp` or `impacttmp`. It is recommended that you change the default to a different job name for each new job. When a job is started, several files are created using the job name as the base name for the file names. If the new job has the same name as the previous job, Maestro does not automatically assign new names to jobs or files; it overwrites the files with data from the new job. Files from the first job will not be recoverable.

Login

To run Impact on a remote machine, you must specify a login name for the remote machine. By default, the login name is set to the login name of the user who began the Maestro session. If a different login name is required, it must be typed here before the job is started. If the Impact job is running on a local computer, the Login field is ignored.

Host

On this menu, Maestro will display all hosts defined in the file `$SCHRODINGER/schrodinger.hosts`.

The following two options appear on the panels for QSite and for Basic Impact applications only.

Source of job input

This feature is not available for Glide and Liaison jobs, in which external structure databases are commonly used. When you set up other FirstDiscovery jobs, you may use this feature to choose either the on-screen Workspace structure or a Selected entry from the Project Table.

The Workspace selection is the default choice, indicating that the simulation to be performed will operate on whatever atoms, molecules, or entries are part of the on-screen structure, even those atoms hidden by the Display/Undisplay Atom facility. Performing a simulation with Frozen or Buffered Atoms (see the Constraints folder, Section 9.5 on page 160) requires that you use the Workspace structure.

The Selected entry option is for running the simulation on whatever entry is currently selected in the Project Table, irrespective of what is displayed in the Workspace. Note that only one entry can be selected. If more than one entry is selected in the project, an error message will appear. This option is incompatible with selecting Frozen or Buffered Atoms constraints.

Incorporate output into project by

When a QSite or Basic Impact job is completed, the structural results are incorporated into the project that was open when the job was started. The choice of option determines how this is done:

- Appending new entries (default): Each structure in the output file is added to the project as a new entry.
- Replacing existing entries: Any entries that served as input for the job are replaced with the new structures returned from the calculation.
- Do not incorporate: No change is made to the project when the job is complete.

1.2.2 Panel Controls and Tabbed Folders

Controls that are specific to a panel appear in the middle section of the panel. Usually the controls will be contained in a series of tabbed folders. Each folder contains settings relevant to the task that the panel performs.

1.2.3 Start, Write, Hide, and Help

All FirstDiscovery panels have the following buttons in the lower portion of the panel:

Start

Submits the job to the specified Host.

Write Job Files

Writes out files required for the job without starting the job. The job can be run from the command line in a UNIX shell.

Hide

Dismisses the current panel without starting the job or writing any files.

Unlike other panels in Maestro, only one Impact panel can be open at a time. If you choose an item from the Impact menu while another panel is open, the existing panel is dismissed before the new one is displayed.

Help

Opens the Help panel with an appropriate help topic displayed.

Chapter 2: Introduction to Maestro

Maestro™ is the graphical user interface for all of Schrödinger's products: FirstDiscovery™ (Glide™, Impact™, Liaison™, and QSite™), Jaguar™, LigPrep™, MacroModel®, Prime™, and QikProp™. It contains tools for building, displaying, and manipulating chemical structures; for organizing, loading, and storing these structures and associated data; and for setting up, monitoring, and visualizing the results of calculations on these structures. This chapter provides a brief introduction to Maestro and some of its capabilities. For more information, see the *Maestro User Manual*.

2.1 General Interface Behavior

Most Maestro panels are amodal: more than one panel can be open at a time, and a panel need not be closed for an action to be carried out. Instead of a Close menu option or button, each Maestro panel has a Hide button so that you can hide the panel from view.

Maestro supports the mouse functions common to many graphical user interfaces. The left button is used for selecting: choosing menu items, clicking buttons, and selecting objects. This button is also used for resizing and moving panels. The right button displays a context-sensitive menu. Other common mouse functions are supported, such as using the mouse in combination with the SHIFT or CTRL keys to select a range of items and select or deselect a single item without affecting other items.

In addition, the mouse buttons are used for special functions described later in this chapter. These functions assume that you have a three-button mouse. If you have a two-button mouse, ensure that it is configured for three-button mouse simulation (the middle mouse button is simulated by pressing or holding down both buttons simultaneously).

2.2 Starting Maestro

Before you start Maestro, you must first set the `SCHRODINGER` environment variable to point to the installation directory. You can set this variable by entering the following command at a shell prompt:

csh/tcsh: `setenv SCHRODINGER` *installation-directory*

bash/ksh: `export SCHRODINGER=`*installation-directory*

You might also need to set the `DISPLAY` environment variable if it is not set automatically when you log in. To determine if you need to set this variable, enter the command:

```
echo $DISPLAY
```

If the response is a blank line, set the variable by entering the following command:

csh/tcsh: `setenv DISPLAY` *display-machine-name*`:0.0`

bash/ksh: `export DISPLAY=`*display-machine-name*`:0.0`

After you set the `SCHRODINGER` and `DISPLAY` environment variables, you can start Maestro using the command:

`$SCHRODINGER/maestro` *options*

If the `$SCHRODINGER` directory has been added to your path, you only need to enter the command `maestro`. Options for this command are given in the *Maestro User Manual*.

The directory from which you started Maestro is Maestro's current working directory, and all data files are written to and read from this directory unless otherwise specified (see Section 2.8 on page 23). You can change directories by entering the following command in the command input area of the main window:

`cd` *directory_name*

where *directory_name* is either a full path or a relative path.

2.3 The Maestro Main Window

The Maestro main window is shown in Figure 2.1 on page 7. The main window components are as follows:

- **Title bar**—displays the project name and the current working directory
- **Auto-Help**—automatically displays context-sensitive help
- **Menu bar**—provides access to panels
- **Workspace**—displays molecular structures
- **Clipping planes window**—displays a small, top view of the Workspace and shows the clipping planes and viewing volume indicators
- **Toolbar**—contains buttons for many common tasks, and also provides tools for displaying and manipulating structures and organizing the Workspace
- **Status bar**—displays the number of atoms, entries, residues, chains, and molecules in the Workspace
- **Sequence viewer**—shows the sequences for proteins displayed in the Workspace
- **Command input area**—provides a place to enter Maestro commands

You can control the display of any of the last five components from the Display menu.

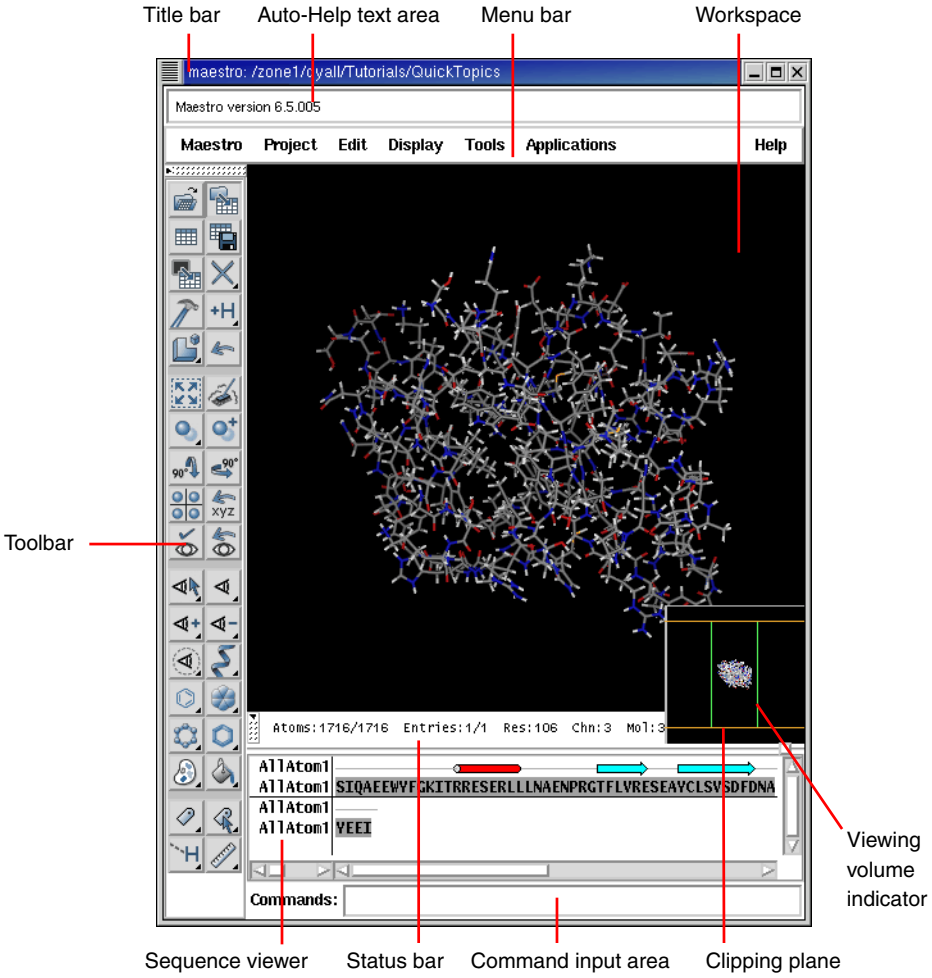


Figure 2.1. The Maestro main window.

When a distinction between components in the main window and those in other panels is needed, the term *main* is applied to the main window components (e.g., main toolbar).

2.3.1 The Menu Bar

The menus on the main menu bar provide access to panels, allow you to execute some commands, and control the appearance of the Workspace. The main menus are as follows:

- **Maestro menu**—allows you to save or print images in the Workspace, execute system commands, save or load a panel layout, set preferences, set up Maestro command aliases, and quit Maestro.

- **Project menu**—provides access to many project-related actions, such as opening and closing a project and importing and exporting structures. These actions can also be performed from the Project Table panel. For more information, see Section 2.4 on page 12.
- **Edit menu**—allows you to undo actions, build and modify structures, define command scripts and macros, and find atoms in the Workspace.
- **Display menu**—provides access to panels that control the display of the contents of the Workspace, to commands for the display of panels, and to options for the display of main window components. Your choice of main window components displayed is persistent between Maestro sessions.
- **Tools menu**—provides access to panels for grouping atoms, for measuring, for aligning and superpositioning structures, and for viewing and visualizing data.
- **Applications menu**—provides access to panels for setting up, submitting, and monitoring jobs for all Schrödinger's computational programs. Some products have a submenu from which you can choose the task to be performed.
- **Help menu**—provides access to the Help panel and to a control for the display of Balloon help (tooltips).

2.3.2 The Toolbar

The main toolbar contains buttons for performing common tasks. There are three kinds of buttons on the toolbar:

- Those that perform simple tasks, like clearing the Workspace
- Those that show and hide panels or open dialog boxes
- Those that display a menu when you click and hold

The third type of button is called a *menu button* and has a triangle in the lower right corner. The menu is called a *button menu*. There are two kinds of items on button menus, and both kinds can be on the same menu:

- Actions, which perform an action immediately
- States, which you set before an action is performed

When you select a state, it is stored and marked with a red diamond on the menu. Most states on button menus are pick states (see Section 2.6 on page 21), which means you must pick an atom in the Workspace before the action is performed. If you click a menu button that has pick states, the button is indented to indicate that picking with the selected pick state is in effect. If you double-click a menu button that has pick states, the action is applied to all atoms. Some other menu buttons support double-clicking to apply an action: this support is explicitly mentioned in the button description.

You can show or hide the toolbar using the collapse button at the top or by selecting Toolbar from the Display menu. You can hide it or move it to the right or left side of the Workspace by right-clicking in the toolbar and selecting the appropriate option.

The buttons are described below. Some descriptions refer to features not described in this chapter. See the *Maestro User Manual* for a fuller description of these features.

Open a project
Open the Open Project dialog box.

Import structures
Show the Import panel.

Show/Hide project table
Show the Project Table panel or hide it if it is displayed.

Save as
Open the Save Project As dialog box, to save the project with a new name.

Create entry from workspace
Create an entry in the current project using the contents of the Workspace.

Delete
Choose an object to delete. Menu button with a pick menu, a section to delete hydrogens and waters and to open the Atom Selection dialog box, and a section to delete other objects associated with the structures in the Workspace.

Show/Hide Build panel
Show the Build panel or hide it if it is displayed.

Add hydrogens
Pick atoms for hydrogen treatment. Menu button with a pick menu and an item to open the Atom Selection dialog box.

Local transformation
Pick the object to transform. Menu button with a pick menu and an item to open the Advanced Transformations panel.

Undo/Redo
Undo or redo the last action. Performs the same function as the Undo item on the Edit menu, and changes to an arrow pointing in the opposite direction when an Undo has been performed, indicating that its next action is Redo.

Fit to screen
Scale what is displayed to fit into the Workspace, and reset the center of rotation.

Clear workspace
Clear all atoms from the Workspace

Set fog display state
Menu button. Automatic means *on* when there are more than 40 atoms in the Workspace, *off* when there are fewer.

Enhance depth cues
Optimize fogging and other depth cues based on what is in the Workspace.

Rotate around X axis by 90 degrees

Rotate around Y axis by 90 degrees

Tile entries
Arrange entries in a rectangular grid in the Workspace.

Reset workspace
Reset the rotation, translation, and zoom of the Workspace to the default state.

Save view
Save the current view of the Workspace: orientation, location, and zoom.

Restore view
Restore the last saved view of the Workspace: orientation, location, and zoom.

Display only picked atoms
Pick atoms to display. Menu button with a pick menu.

Display only
Display only the selected atoms. Menu button with a list of predefined atom categories and an item to open the Atom Selection dialog box.

Also display
Add the selected atoms to the display. Menu button with a list of predefined atom categories and an item to open the Atom Selection dialog box.

Undisplay
Undisplay the selected atoms. Menu button with a list of predefined atom categories and an item to open the Atom Selection dialog box.

Display residues within N angstroms of currently displayed atoms
Menu button with a list of values and an item to open a dialog box to set a value.

Show, hide, or color ribbons
Menu button with items to control the display of ribbons and atoms for proteins and to color ribbons by various schemes.

Draw bonds in wire
Pick atoms for representation. Menu button with a pick menu and an item to open the Atom Selection dialog box.

Draw atoms in CPK
Pick atoms for representation. Menu button with a pick menu and an item to open the Atom Selection dialog box.

Draw atoms in ball and stick
Pick atoms for representation. Menu button with a pick menu and an item to open the Atom Selection dialog box.

Draw bonds in tube
Pick atoms for representation. Menu button with a pick menu and an item to open the Atom Selection dialog box.

Color all atoms by scheme
Menu button with a list of schemes.

Color residue by constant color
Pick residues to apply the selected color. Double-click to color all atoms. Menu button with a list of colors.

Label atoms
Label all atoms with the selected label. Menu button with a list of label types and an item to delete labels.

Label picked atoms
Menu button with a pick menu and items to open the Atom Selection dialog box, to open the Atom Labels panel at the Composition folder, and to delete labels.

Display H-bonds
Pick molecules to display H-bonds. Menu button with items to choose to display H-bonds within the selected molecule (intra) or between the selected molecule and all other atoms in the Workspace (inter), or to delete H-bonds.

Measure distances, angles or dihedrals
Pick atoms to define measurements. Menu button with items to choose between distance (default), angle, or dihedral measurement, and to delete measurements.

2.3.3 Mouse Functions in the Workspace

The middle and right mouse buttons have special uses in the Workspace. These can be used on their own and in combination with the SHIFT and CTRL keys to perform common operations, such as rotation, translation, centering, and zooming. Apart from centering a molecule on an atom, all these operations involve dragging.

Table 2.1. Mapping of Workspace operations to mouse actions.

Operation	Action
Rotate about the *x*- and *y*-axes	Drag with middle mouse button
Rotate about the *x*-axis only	Drag vertically with SHIFT and middle mouse button
Rotate about the *y*-axis only	Drag horizontally with SHIFT and middle mouse button
Rotate about the *z*-axis	Drag horizontally with CTRL and middle mouse button
Spot-center on an atom	Right-click
Translate in the *x*-*y* plane	Drag with right mouse button
Translate along the *y*-axis	Drag vertically with SHIFT and right mouse button
Translate along the *x*-axis	Drag horizontally with SHIFT and right mouse button
Translate about the *z*-axis	Drag horizontally with CTRL and right mouse button
Zoom	Drag horizontally with middle and right mouse buttons or with SHIFT+CTRL and middle mouse button

2.3.4 Shortcut Key Combinations

Some frequently used operations that can be performed in the main window have been assigned shortcut key combinations. The shortcuts, their functions, and their menu equivalents are listed in Table 2.2.

Table 2.2. Shortcut keys in the Maestro main window.

Keys	Action	Equivalent Menu Choices
ALT+B	Show Build panel	Edit > Build
ALT+C	Create entry	Project > Create Entry From Workspace
ALT+E	Show Command Script Editor panel	Edit > Command Script Editor
ALT+F	Open Find Atoms panel	Edit > Find
ALT+H	Show Help panel	Help > Help
ALT+I	Show Import panel	Project > Import Structures
ALT+M	Show Measurements panel	Tools > Measurements
ALT+N	New project	Project > New
ALT+O	Open project	Project > Open
ALT+P	Print	Maestro > Print
ALT+Q	Quit	Maestro > Quit
ALT+S	Show Sets panel	Tools > Sets
ALT+T	Show Project Table panel	Project > Show Table
ALT+W	Close project	Project > Close
ALT+Z	Undo/Redo last command	Edit > Undo/Redo

2.4 Maestro Projects

All the work you do in Maestro is done within a *project*. A project consists of a set of *entries*, each of which contains one or more chemical structures and their associated data. In any Maestro session, there can be only one Maestro project open. If you do not specify a project when you start Maestro, a *scratch* project is created. You can work in a scratch project without saving it, but you must save it in order to use it in future sessions. Maestro also creates a scratch project when you close a project.

Likewise, if there is no entry displayed in the Workspace, Maestro creates a *scratch* entry. Structures that you build in the Workspace constitute a scratch entry until you save the structures as project entries. The scratch entry is not saved with the project unless you explicitly incorporate it into the project. However, you can use a scratch entry as input for some calculations.

The structures and their data are represented in the Project Table, which displays a list of entries. Each entry is represented by a row in the Project Table. Each row contains the row number, an icon indicating whether the entry is displayed in the Workspace (the In

column), the entry title, a button to open the Surfaces panel if the entry has surfaces, the entry name, and any entry properties. The row number is not a property of the entry.

You can open the Project Table panel by choosing Show Table from the Project menu, by clicking the Show/Hide project table button on the toolbar, or by pressing ALT+T.

The Project Table panel contains a menu bar, a toolbar, and the table itself.

You can use entries as input for all of the computational programs—Glide, Impact, Jaguar, Liaison, LigPrep, MacroModel, Prime, QikProp, and QSite. You can select entries as input for the ePlayer, which displays the selected structures in sequence. You can also duplicate, combine, rename, and sort entries, create properties, import structures as entries, and export structures and properties from entries in various formats.

2.4.1 The Project Table Toolbar

The Project Table toolbar contains two groups of buttons and a status display. The first set of buttons opens various panels that allow you to perform functions on the entries in the Project Table. The second set of buttons controls the ePlayer, which "plays through" the selected structures: each structure is displayed in the Workspace in sequence, at a given time interval. See Section 2.3.2 on page 8 for a description of the types of toolbar buttons. The buttons are described below.

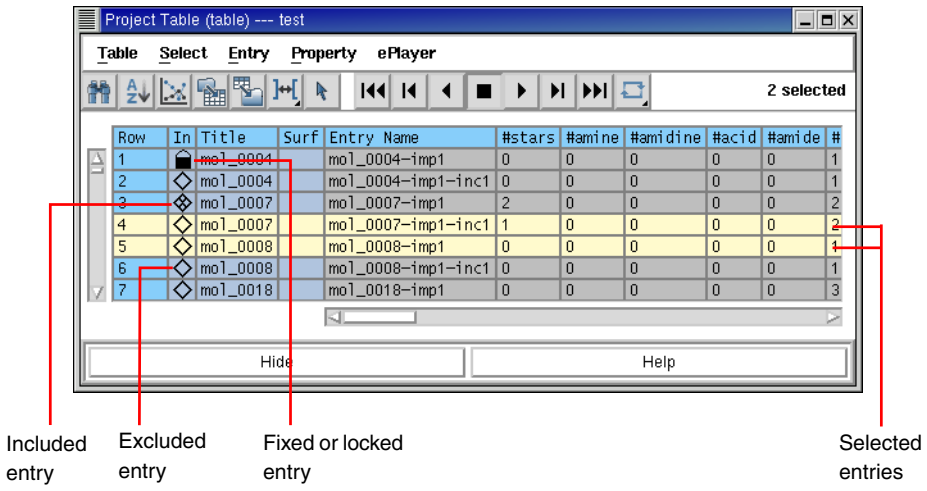

Project Table (table) --- test
Table Select Entry Property ePlayer
2 selected
Row In Title Surf Entry Name #stars #amine #amidine #acid #amide #
1 mol_0004 mol_0004-imp1 0 0 0 0 0 1
2 mol_0004 mol_0004-imp1-inc1 0 0 0 0 0 1
3 mol_0007 mol_0007-imp1 2 0 0 0 0 2
4 mol_0007 mol_0007-imp1-inc1 1 0 0 0 0 2
5 mol_0008 mol_0008-imp1 0 0 0 0 0 1
6 mol_0008 mol_0008-imp1-inc1 0 0 0 0 0 1
7 mol_0018 mol_0018-imp1 0 0 0 0 0 3
Hide
Help
Included entry
Excluded entry
Fixed or locked entry
Selected entries

Figure 2.2. The** Project Table **panel.

Find
Open the Find panel for locating alphanumeric text in any column of the Project Table, except for the row number.

Sort
Open the Sort panel for sorting entries by up to three properties.

Plot
Open the Plot panel for plotting entry properties.

Import structures
Open the Import panel for importing structures into the project.

Export structures
Open the Export panel for exporting structures to a file.

Columns menu
Display a menu for adjusting the column widths.

Entry selection
Open the Entry Selection dialog box for selecting entries based on criteria for entry properties.

Go to start
Display the first selected structure.

Previous
Display the previous structure in the list of selected structures.

Play backward
Display the selected structures in sequence, moving toward the first.

Stop
Stop the ePlayer.

Play forward
Display the selected structures in sequence, moving toward the last.

Next
Display the next structure in the list of selected structures.

Go to end
Display the last selected structure.

ePlayer loop menu
Display a menu of options for repeating the display of the structures. Single direction displays structures in a single direction, then repeats. Oscillate reverses direction each time the beginning or end of the list is reached.

The status display shows the number of selected entries. When you pause the cursor over the display, the Balloon help shows the total number of entries, the number shown in the table, the number selected, and the number included.

2.4.2 The Project Table Menus

- **Table menu**—provides tools for finding text, sorting entries, and plotting properties, importing and exporting structures, and configuring the Project Table.
- **Select menu**—provides commands and access to the Entry Selection dialog box and the Filter panel so that you can select entries.
- **Entry menu**—provides tools for including and excluding entries, controlling the display of entries in the Project Table, and performing various operations on the selected entries.
- **Property menu**—provides tools for displaying and manipulating properties.
- **ePlayer menu**—provides access to the play controls and the ePlayer options.

2.4.3 Selecting Entries

Many operations in Maestro are performed on the entries that are selected in the Project Table. The Project Table functions much like any other table: you select rows by clicking, shift-clicking, and control-clicking. However, because clicking in an editable cell of a selected row enters edit mode, you should click in the Row column to select entries. See Section 2.4.5 on page 16 for more information on mouse actions. There are shortcuts for selecting classes of entries on the Select menu.

In addition to selecting entries manually, you can select entries that meet a combination of conditions on their properties. Such combinations of conditions are called *filters*. Filters are Entry Selection Language (ESL) expressions and are evaluated at the time they are applied. For example, if you want to set up a Glide job that uses ligands with a low molecular weight (say, less than 300) and that has certain QikProp properties, you can set up a filter and use it to select entries for the job. If you save it, you can use it again on a different set of ligands that met the same selection criteria.

You can create filters in the Entry Selection dialog box, which you can open from the Select menu (Only, Add, Deselect), from the Edit Filter dialog box, or by clicking the Entry selection button on the toolbar.

To create a filter, choose a property from the property list, choose a condition, and combine it with the current filter by clicking Add, Subtract, or Intersect. These buttons perform the Boolean operations OR, AND NOT, and AND on the corresponding ESL expressions. Once you have created a filter, you can click OK to apply it immediately or name and save it for later use.

2.4.4 Including Entries in the Workspace

In addition to selecting entries for various tasks, you also control which entries are displayed in the Workspace from the Project Table. An entry that is displayed in the Workspace is said to be *included* in the Workspace; likewise, an entry that is not displayed is *excluded*. Included entries are marked by an X in the diamond in the In column; excluded entries are marked by an empty diamond. Entry inclusion is completely independent of entry selection.

To include or exclude entries, you can click, shift-click, and control-click in the In column, or select entries and then include or exclude them from the Entry menu. Inclusion with the mouse works just like selection: when you include an entry by clicking, all other entries are excluded.

It is sometimes useful to keep one entry in the Workspace and include others one by one: for example, a receptor and a set of ligands. You can fix entries in the Workspace by selecting the entries and choosing Fix from the Entry menu or by pressing ALT+F. A padlock icon replaces the diamond in the In column to denote a *fixed* entry. To remove a fixed entry from the Workspace, you must exclude it explicitly (ALT+X). It is not affected by the inclusion or exclusion of other entries. Fixing affects only the inclusion of the entry: you can still rotate, translate, or modify the structure.

2.4.5 Mouse Functions in the Project Table

The Project Table supports the standard use of shift-click and control-click to select objects. This behavior applies to the selection of entries and the inclusion of entries in the Workspace. Dragging to resize rows and columns and to move rows is also supported.

You can drag a set of non-contiguous entries to reposition them in the Project Table. When you release the mouse button, the entries are placed after the first unselected entry that precedes the entry on which the cursor is resting. For example, if you select entries 2, 4, and 6, and release the mouse button on entry 3, these three entries are placed after entry 1, because entry 1 is the first unselected entry that precedes entry 3. To move entries to the top of the table, drag them above the top of the table; to move entries to the end of the table, drag them below the end of the table.

A summary of project-based mouse functions is provided in Table 2.3.

Table 2.3. Mouse operations in the Project Table.

Task	Mouse Operation
Change a Boolean property value	Click repeatedly in a cell to cycle through the possible values (On, Off, Clear)
Display the Entry menu for an entry	Right-click anywhere in the entry. If the entry is not selected, it becomes the selected entry. If the entry is selected, the action is applied to all selected entries.
Display a version of the Property menu for a property	Right-click in the column header
Edit the text or the value in a table cell	Click in the cell and edit the text or value
Include an entry, exclude all others	Click the In column of the entry
Move selected entries	Drag the entries
Paste text into a table cell	Middle-click
Resize rows or columns	Drag the boundary with the middle mouse button
Select an entry, deselect all others	For an unselected entry, click anywhere in the row except the In column; for a selected entry, click the row number
Select or include multiple entries	Click the first entry then shift-click the last entry
Toggle the entry selection or inclusion state	Control-click the entry or the In column

2.4.6 Project Table Shortcut Keys

Some frequently used project operations have been assigned shortcut key combinations. The shortcuts, their functions, and their menu equivalents are listed in Table 2.4.

Table 2.4. Shortcut keys in the Project Table.

Keys	Action	Equivalent Menu Choices
ALT+A	Select all entries	Select > All
ALT+F	Fix entry in Workspace	Entry > Fix
ALT+I	Show import panel	Table > Import Structures
ALT+N	Include only selected entries	Entry > Include Only
ALT+U	Deselect all entries	Select > None
ALT+X	Exclude selected entries	Entry > Exclude
ALT+Z	Undo/Redo last command	Edit > Undo/Redo in main window

2.5 Building a Structure

After you start Maestro, the first task is usually to create or import a structure. You can open existing Maestro projects or import structures from other sources to obtain a structure. To build a structure, you use the Build panel, which you can open by choosing Fragments from the Edit menu, or by clicking the Show/Hide Build panel button in the toolbar.

The Build panel allows you to create structures by drawing or placing atoms or fragments in the Workspace, and connecting them into a larger structure, to adjust atom positions and bond orders, and to change atom properties. This panel contains a toolbar and three folders.

The Fragments folder offers a variety of molecular fragments from which to build a structure.

To place a fragment in the Workspace:

1. Select Place.
2. Choose a fragment library from the Fragments menu.
3. Click a fragment.
4. Click in the Workspace where you want the fragment to be placed.

There are several options for adding to a fragment that you have placed:

- Place another fragment and connect them using the Connect & Fuse panel, which you open from the Edit menu on the main menu bar or with the Display Connect/Fuse panel on the Build toolbar.

- Replace one or more atoms in the existing fragment with another fragment by selecting a fragment and clicking in the Workspace on the main atom to be replaced.
- Grow another fragment by selecting Grow and clicking the fragment you want to add in the Fragments folder.

  Grow mode uses predefined rules to connect a fragment to the *grow bond*. The grow bond is marked by a green arrow. The new fragment replaces the atom at the head of the arrow on the grow bond and all atoms attached to it. You can change the grow bond by clicking on the desired grow bond in the Workspace. The arrow points to the atom nearest to where you clicked.

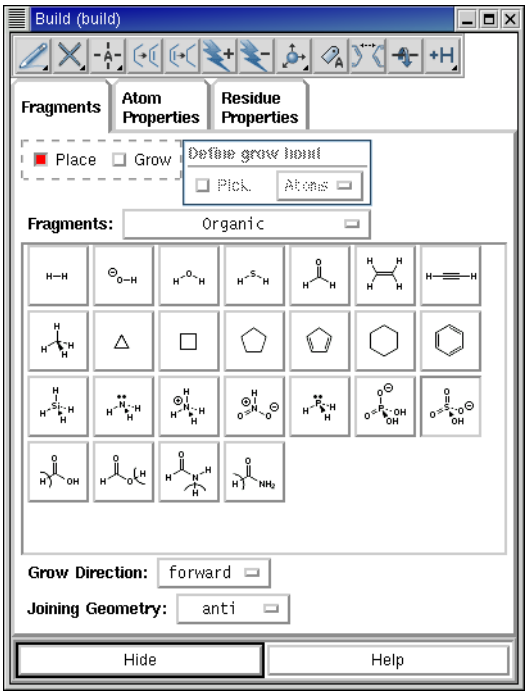


Figure 2.3. The* Build *panel.

You can also draw a structure freehand by choosing an element from the Draw button menu on the Build panel toolbar and then drawing the structure.

In the Atom Properties folder you can change the properties of the atoms in the Workspace. For each item on the Property menu—Element, MacroModel Type, Partial Charge, PDB Atom Name, Grow Name, and Atom Name—there is a set of tools you can use to change the atom properties. For example, the Element tools consist of a periodic table from which you can choose an element and pick an atom to change it to an atom of the selected element.

Similarly, the Residue Properties folder provides tools for changing the properties of residues: the residue number, the residue name, and the chain name.

To adjust bond lengths, bond angles, dihedral angles, and chiralities during or after building a structure, use the Adjust panel, which you open from the Edit menu on the main menu bar or with the Display Adjust panel button on the Build panel toolbar.

The toolbar of the Build panel provides quick access to tools for drawing and modifying structures and labeling atoms. See Section 2.3.2 on page 8 for a description of the types of toolbar buttons. The toolbar buttons and their use are described below.

Draw
Draw structures freehand in the Workspace. Menu button with a list of elements to draw with (default C). Each click in the Workspace places an atom and connects it to the previous atom.

Delete
Choose an object to delete. Menu button with a pick menu and other items. Same as the Delete button on the main toolbar.

Set element
Pick atoms to change to the selected element (default C). Menu button with a list of target elements. Menu button with a limited list of target elements.

Increment bond order
Pick a bond to increase its bond order by one, to a maximum of 3.

Decrement bond order
Pick a bond to decrease its bond order by one, to a minimum of 0.

Increment formal charge
Pick an atom to increase its formal charge by one.

Decrement formal charge
Pick an atom to decrease its formal charge by one.

Move
Pick an atom to move in the xy plane or the z direction. Menu button with a list of directions. Moves in the xy plane are made by clicking the new location. Moves in the z direction are made in 0.5 Å increments.

Label
Apply heteroatom labels as you build a structure. The label consists of the element name and formal charge, and is applied to atoms other than C and H.

Display Connect/Fuse panel
Open the Connect & Fuse panel so you can connect structures (create bonds between structures) or fuse structures (replace atoms of one structure with those of another).

Display Adjust panel
Open the Adjust panel so you can change bond lengths, bond angles, dihedral angles, or atom chiralities.

Add hydrogens
Pick to apply the current hydrogen treatment. Menu button with a pick menu and an item to open the Atom Selection dialog box. Same as the Add hydrogens button on the main toolbar.

2.6 Selecting Atoms

Maestro has a powerful set of tools for selecting atoms in a structure that takes advantage of chemical information about the structure. These tools are embedded in each panel in which you might need to select atoms to apply some operation. Once you have chosen an operation, you can use the tools to select, or *pick,* the atoms to which to apply the operation.

To select all atoms in a molecule, a chain, a residue, or an entry, you can choose a *pick state* using the Pick menu. Once you have chosen the pick state, you can click on an atom in the Workspace, and all the atoms that belong to the same structural unit, as defined by the pick state, are selected. For example, if you choose Residue and click on any atom in a glycine residue, all the atoms in that glycine residue are selected. To select individual atoms, choose Atoms from the Pick menu. The Pick menu varies from panel to panel, because not all pick states are appropriate for a given operation. For example, some panels have only Atoms and Bonds in the Pick menu.

To make atom selections based on more complex criteria, such as all carbon atoms in a protein backbone, you can use the Atom Selection dialog box. To open this dialog box, click the Select button. You can select an atom group from any of the folders in the dialog

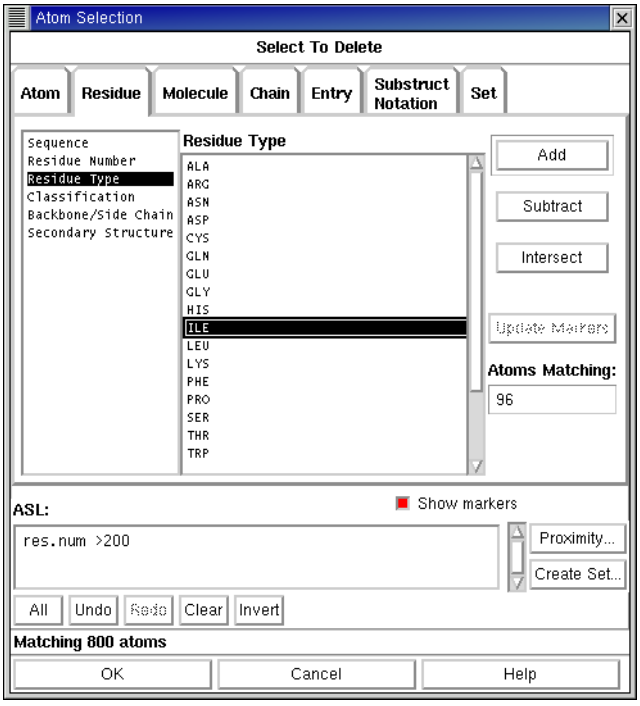


Figure 2.4. The Atom Selection ***dialog box.***

box: Atom, Residue, Molecule, Chain, Entry, Substruct Notation, or Set. You can then combine this group with the existing atom group using the buttons on the right: the Add button (Boolean OR) includes all atoms in the new group or the existing group; the Subtract button (Boolean AND NOT) excludes atoms in the new group from the existing group; and the Intersect button (Boolean AND) includes only those atoms that are in both the new group and the existing group. The existing group is expressed in Atom Specification Language (ASL) in the ASL text box, and is shown with light blue markers in the Workspace. The current selection is shown with purple markers.

When you are satisfied with the selection, click OK to apply the operation you have chosen to the selection you have made. The operation is described in a bar at the top of the Atom Selection dialog box. Some operations take effect immediately, such as deleting atoms. Others merely define a set of atoms to be used in a subsequent task, such as selecting atoms for the creation of a surface.

When the Atom Selection dialog box is open, you cannot perform other actions except for rotation and translation of structures and picking. You can also open the help viewer.

2.7 Maestro Command Scripts

Although you can perform nearly all Maestro-supported operations through menus and panels, you can also perform operations using Maestro commands, or compilations of these commands, called *scripts*.

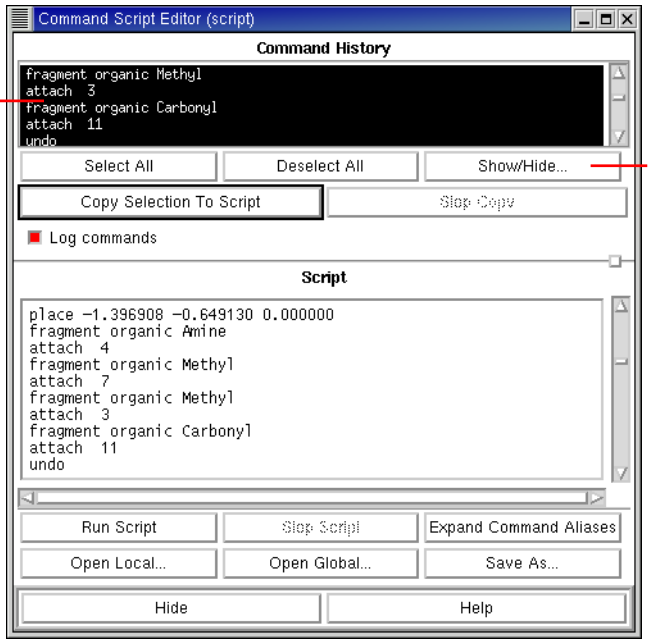
Command scripts can be used to automate lengthy procedures or repetitive tasks. Because all Maestro commands are logged and displayed in the Command Script Editor panel, you can create a command script by performing the operations with the GUI controls, copying the logged commands from the Command History list into the Script area of the panel, then saving the list of copied commands as a script. Short scripts can also be saved as macros, which are run from the keys F1 through F12. See the *Maestro User Manual* for details.

To run an existing command script:

1. Open the Command Script Editor panel from the Edit menu in the main window.
2. Click Open Local and navigate to the directory containing the desired script.
3. Select a script in the Files list and click Open.

   The command script is loaded into the Script window of the Command Script Editor panel.
4. Click Run Script.

Command scripts cannot be used for Prime operations.



The *Command History* window displays a log of all commands issued internally within Maestro when you interact with a panel, menu, or structure

Opens the *Show/Hide Command* panel, used to determine which commands are logged in the *Command History* list

Figure 2.5. The Command Script Editor ***panel.***

2.8 Specifying a Maestro Working Directory

When you use Maestro to launch FirstDiscovery jobs, Maestro writes job output to the directory specified in the Directory folder of the Preferences panel. By default, the directory to which Maestro writes files (the file I/O directory) is the directory from which you started Maestro.

To change this directory:

1. Open the Preferences panel from the Maestro menu.
2. Click the Directory tab.
3. Select the option for the directory you want files to be read from and written to.

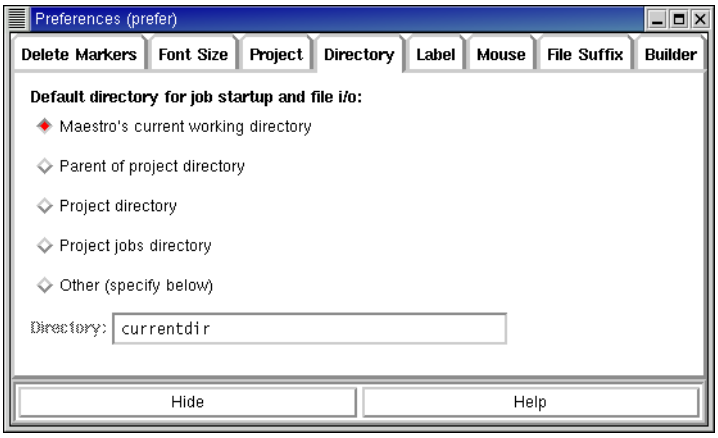


Figure 2.6. The** Directory **folder of the** Preferences **panel.

2.9 Undoing an Operation

To undo a single operation, click the Undo button in the toolbar, choose Undo from the Edit menu, or press ALT+Z. The word Undo in the menu is followed by text that describes the operation or operations to undo. Not all operations can be undone: for example, global rotations and translations are not undoable operations. For such operations you can use the Save view and Restore view buttons in the toolbar, which save and restore a molecular orientation.

If you think that you might want to undo a series of operations later, you can start an undo block by selecting Begin Undo Block from the Edit menu. When you have completed the group of operations you want to undo, end the block by selecting End Undo Block from the Edit menu. Then, to undo the operations in the block, choose Undo from the Edit menu. Undo is not supported for all Maestro operations. An undo block can be created only if at least one undoable operation has been performed since the Begin Undo Block command was issued.

2.10 Running and Monitoring Jobs

While FirstDiscovery jobs can be run from the command line, we suggest that you use the Maestro GUI to set up and launch these jobs, at least until you have some experience with the programs and understand the directory structure and the input file requirements. Maestro has dedicated panels for each product for preparing and submitting jobs. To use these panels, make the appropriate choice for the product and task from the Applications menu and its submenus.

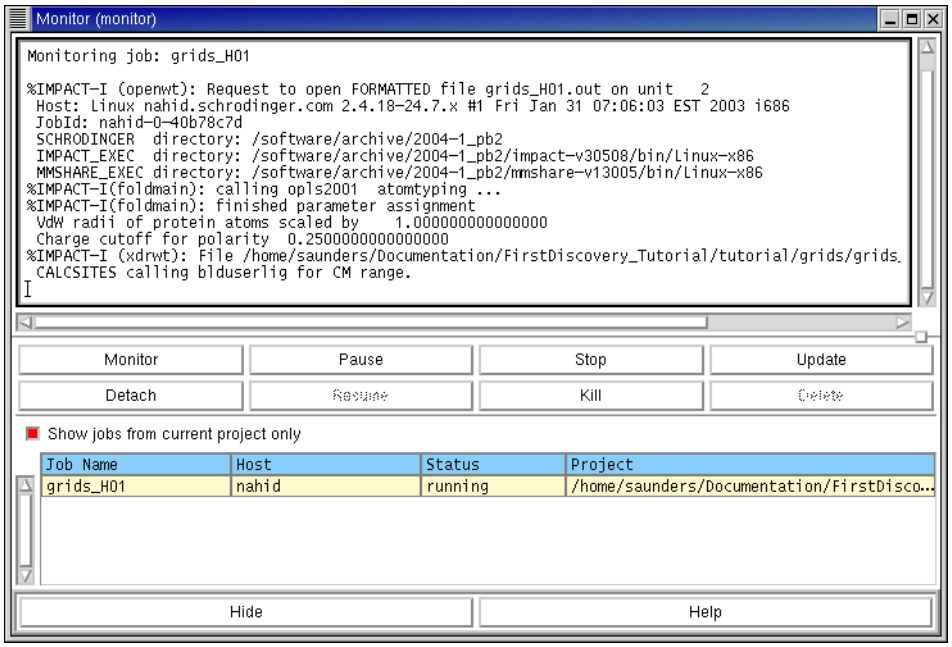


Figure 2.7. The* Monitor *panel.

Maestro also has a job control panel for monitoring the progress of jobs and for pausing, resuming, or killing jobs—the Monitor panel. All jobs that belong to your user ID can be displayed in the Monitor panel, whether or not they were started from Maestro. The text pane shows some kinds of output from the job that is being monitored, such as the contents of the log file. The Monitor panel opens automatically when you start a job. If it is not open, you can open it by choosing Monitor from the Applications menu in the Maestro main window.

While jobs are running, the Detach, Pause, Resume, Stop, Kill, and Update buttons are active. When there are no jobs currently running, only the Monitor and Delete buttons are active. These buttons act on the selected job. By default, only jobs started from the current project are shown. To show other jobs, deselect Show jobs from current project only.

When a job that is being monitored ends, the results are automatically incorporated into the project. If a job that is not currently being monitored ends, you can select it in the Monitor panel and incorporate the results. Monitored jobs are incorporated only if they are part of the project. You can monitor jobs that are not part of the project, but their results are not incorporated. To add their results to the project, you must import them.

2.11 Getting Help

Maestro comes with automatic, context-sensitive help (Auto-Help), Balloon help (tooltips), an online help facility, and a user manual. To get help, follow the steps below:

- Check the Auto-Help text box below the title bar of the main window. If help is available for the task you are performing, it is automatically displayed there. It describes what actions are needed to perform the task.
- If your question concerns a GUI element, such as a button or option, there may be Balloon help for the item. Pause the cursor over the element. If the Balloon help does not appear, check that Show Balloon Help is selected in the Help menu of the main window. If there is Balloon help for the element, it appears within a few seconds.
- If you do not find the help you need using either of the steps above, click the Help button in the lower right corner of the appropriate panel. The Help panel is displayed with a relevant help topic.
- For help with a concept or action not associated with a panel, open the Help panel from the Help menu or press ALT+H.

If you do not find the information you need in the Maestro help system, check the following sources:

- The *Maestro User Manual*
- The *Maestro Release Notes*
- The Frequently Asked Questions page, found at http://www.schrodinger.com/Support/faq.html

2.12 Ending a Maestro Session

To end a Maestro session, choose Quit from the Maestro menu. To save a log file with a record of all operations performed in the current session, click Quit, save log file in the Quit panel. This information can be useful to Schrödinger support staff when responding to any problem you report.

Chapter 3: FirstDiscovery from the Command Line

3.1 Command-Line Capabilities

This section outlines the capabilities of the FirstDiscovery suite as run from the command line, rather than the Maestro GUI.

3.1.1 Location of Files and Working Directory

For both the FirstDiscovery and Basic Impact applications, Maestro normally writes input files to the directory from which you launched Maestro (called the *Maestro working directory*). Impact also normally writes its output files to the same location, though Impact input files and the Glide interface allow you to specify an arbitrary location for grid files. The exception is Liaison, where a directory hierarchy is created based on job names and on the names you assign to the individual ligands being simulated. For more information, see Figure 7.1 on page 115 and Section 7.4 on page 119.

3.1.2 FirstDiscovery Job Types

There are four types of FirstDiscovery jobs that can be run using the `impact` command: Glide, Liaison, QSite, and Basic Impact. The procedure for starting Impact calculations from the command line varies depending on the job type. Follow the specific guidelines listed in this chapter to ensure correct job performance.

3.1.3 Reasons to Run Jobs from the Command Line

Although you will normally set up FirstDiscovery jobs using the controls and settings in the Maestro GUI, you can submit jobs either from within Maestro or from the command line. You sometimes might want to submit jobs from the command line for the following reasons:

- The command-line scripts can run all full-featured jobs written using the FirstDiscovery and Impact panels in Maestro, and also allow you to override specific runtime values that are not accessible through the Maestro interface.
- Command-line scripts allow you to run FirstDiscovery jobs when you want.
- Command-line scripts can be modified and jobs can be re-run without reconfiguring and reloading job settings in Maestro.

- Some job options, such as trajectory file analysis, are available only when you run Impact from the command line.

The Write Job Files button in various Maestro panels writes the input files needed for a job. See Section 3.4 on page 31 for more information.

3.1.4 Force Fields and Write Template

The molecular mechanics force field for Basic Impact applications (and for Glide and Liaison) in FirstDiscovery 3.0 is the OPLS2001 version of OPLS-AA. OPLS2001 is designed to work with automatic atom-typing, and is incompatible with template mode. If you attempt to write a template file (Impact command `WRITE TEMPLATE`) while using OPLS2001, an error message appears to remind you that this command can only be used with OPLS1999 or OPLS2000 force fields. To use one of these older force fields, add a line to your input file before the `CREATE` task, for example:

```
SET FFIELD OPLS1999
```

3.2 File Name Conventions

A typical FirstDiscovery job has one command-script file (*jobname*`.inp`), one or more structure files (*jobname*`.mae`, *jobname*`.pdb`, or *jobname*`.sdf`), and after execution, several output files (e.g., *jobname*`_out.mae` for structure files and *jobname*`.out` for textual data).

If a file already has the name of an output file, in many cases Impact will rename the old file with a numerical extension (*filename*`.out.01`, *filename*`.out.02`, and so on) for archival purposes. The new job's output is then written to the base name (*filename*`.out`). If you do not need the old files, you can remove them.

Some files, such as *jobname*`.log` files, are newly written each time Impact runs a calculation. Likewise, old *jobname*`_pv.mae` files are overwritten. Other examples of files that are *not* incremented are:

- *jobname*`_out.mae` structure files, for Basic Impact minimization and QSite jobs.
- *jobname*`_lig_min.mae` and *jobname*`_rec_min.mae` files, for the minimization section in Liaison.
- *jobname*`_rec_fin.mae` and *jobname*`_lig_fin.mae` files, for the dynamics and HMC sampling methods in Liaison.

In addition, *jobname*`_out.mae` files are *not* produced by default for:

- Liaison jobs—*jobname*`_min.mae` and/or *jobname*`_fin.mae` files are written instead.

- Glide jobs—Glide writes intermediate Maestro-format structure output to *jobname*_raw.mae files, which are incremented.

Table 3.1 contains descriptions of the various file types. For more information, see the Maestro online help or the *FirstDiscovery Command Reference Manual*.

Table 3.1. FirstDiscovery File Extensions.

Extension	Description
.inp	Impact input file or script. Impact input files are formatted plain-text files written in the Impact input file language, DICE. Maestro creates Impact input files before job submission, or you can create or edit them manually with a text editor.
.mae	A Maestro format structure file, a plain-text file written by Maestro containing atom, bond, and other information for one or more molecules.
.log	An Impact log file. If specified, a .log file captures standard output and standard error messages in text form. This file is overwritten during subsequent runs.
.jaguar.in	The Jaguar input file for a QSite calculation.
.out	An Impact output file containing information similar to that found in log files (no standard error). Output files are appended with numerical extensions when the input file is run again. Up to 99 output files are retained.
.01, .02, etc.	A file containing results from previous Impact calculations run from the corresponding *jobname*.inp file.
_out.mae	An Impact output structure file written in the Maestro file format. Glide, Liaison, and some other Impact jobs do not write *_out.mae output structure files.
_raw.mae	Glide's intermediate structure files.

3.3 The Impact Command: Usage Summary

The options that you can specify when initiating jobs from the command line are described in Table 3.2 and Table 3.3. To view the following usage summary information, define the SCHRODINGER environment variable and enter $SCHRODINGER/impact -h in a command shell.

Syntax:

```
impact [options] [[-i] input-file]
```

Table 3.2. `Impact` *Command Options.*

Option	Description
`-h`	Prints usage summary and exits
`-v`	Prints version number of startup script and exits
`-i` *input-file*	Impact input file, conventionally ending in `.inp`. If the *input-file* argument does not end in `.inp`, Impact looks first for *input-file* as specified. If that doesn't exist, it then looks for *input-file*`.inp`. The switch for this argument, `-i`, is optional, but if `-i` is omitted, then *input-file* **must** end in `.inp` and must be the last argument in the command line.
`-o` *output-file*	File for writing output and log messages. If the `-o` option is omitted, Impact will name the log file *jobname*`.log`, where *jobname* is taken from the Impact input file name.
`-s` *size*	Use specific "size" version of the Impact executable. Acceptable values for the `-s` parameter are `medium` or `huge`. If omitted, medium is assumed in most cases; it is valid for up to 8000 atoms or 8000 bonds.
Liaison-Only Options	
`-liasim [-d]` *dir*	Run a Liaison simulation job using ligand directory *dir*. If the `-d` switch is omitted, *dir* must be the last argument on the command line.
`-c` *controlfile*	Specifies name of control file for fit/predict jobs
`-l` *datafile*	Specifies name of data file for fit/predict jobs
`-n` *jobname*	Specifies optional name to use for fit/predict jobs
`-x` *outfile*	Specifies optional output file name for fit/predict jobs
QSite-Only Options	
`-j` *jaguar-file*	Specifies the Jaguar input file
`-p` *num-proc*	Specifies the number of processors to use with Jaguar

Table 3.3. Schrödinger Job Control Options.

Option	Description
`-HOST` *host* `-HOST` *host:n* `-HOST "`*host1 host2*`"`	Specify a remote machine (optionally, its number of processors *n*) on which to run an Impact job. Can also be used to specify a batch queue to submit the job to, or a collection of hosts for distributed or parallel jobs. Default is to run on the local host. See Section 3.4.
`-USER` *user*	Specify a remote user name to run Impact job under. Default is to use the local user name.

Table 3.3. Schrödinger Job Control Options. (Continued)

Option	Description
`-WAIT`	Keep the Impact process in the foreground, instead of returning a command prompt. Does not return until job completes. This is useful in command scripts in which you have specified actions to take only after the Impact job finishes. Without this switch, the Impact job is automatically backgrounded.
`-WHICH`	This switch is a diagnostic tool printing the available Impact installations you can use for the local machine. The job itself is not submitted. The first one listed is the default path; the options `-REL`, `-VER`, and `-ARCH` can direct your job to use a different installation.
`-REL` *release*	This option selects a specific version number of Impact to use. The default is the latest (highest number). Formats like `-REL v3.0`, `-REL 10003`, and `-REL 27` are supported.
`-VER` *pattern*	If you have multiple installations installed, you can specify a pattern with the `-VER` option that matches the installation path to use for your job. The default installation is the one printed by `-WHICH`.
`-ARCH` *platform*	If you have more than one architecture installed for a given system, e.g., AIX-com and AIX-pwr3, then this flag can be used to select either of them, such as `-ARCH pwr3`.
`-LOCAL`	Force remote jobs to run in a local directory, rather than on the remote host. Only active when `-HOST` is used.

3.4 Running Jobs From the Command Line

The `SCHRODINGER` environment variable must be set for Maestro to load and start from a terminal window. You can define `SCHRODINGER` as follows:

csh/tcsh: `setenv SCHRODINGER` *installation-directory*

bash/ksh: `export SCHRODINGER=`*installation-directory*

Entering `ls $SCHRODINGER` at the command prompt will list the Schrödinger installation directory contents, including the Maestro startup script (`maestro`) and the Impact startup script (`impact`).

Unless otherwise specified, Schrödinger applications and utilities run under a job control system and are automatically backgrounded. You need not add an `&` at the end of the commands to have them run and immediately return your command prompt. The `-WAIT` option of the `impact` command prevents automatic backgrounding, so you can embed such commands in other scripts.

`impact -s` *size*

If this option is unspecified, Impact guesses which executable to use based on molecular size values given in the command input file. As a rule of thumb, the `-s huge` option should be used on systems of greater than 8000 atoms or 8000 bonds. However, even on large systems, Glide jobs that dock ligands using previously written grid files, rather than computing grids from a receptor structure, do not require the `-s huge` option.

Note: the `-s huge` option cannot be used with QSite jobs.

3.4.1 Protein Preparation

Protein preparation jobs can be run from the command line using the `protprep` application:

`$SCHRODINGER/protprep` [*options*] *input-file*

or the `pprep` and `impref` utilities. See Section 4.11 on page 55 for more information about command-line protein preparation.

3.4.2 Glide

Using Maestro is the best way to write Glide job scripts, even if you intend to run them from the command line. The scripts are very intricate and are subject to change with each new distribution of the program. To run a Glide job script, enter:

`$SCHRODINGER/impact -i` *jobname*`.inp [-o` *logfile*`]`

By default, the log information is written to *jobname*`.log`, but if you want to use a different file name, use the `-o` option.

If your protein has more than 8000 atoms and you are making grid files (but not if you are just docking ligands), you will also need to include the `-s huge` switch in the command.

The `para_glide` utility can be used to launch a large Glide docking job, distributing the ligand database over a number of processors. See Section 6.11 on page 107 for more information.

While Maestro provides the Pose Viewer interface to visualize high-scoring poses, you can also see the numerical results for these poses in the *jobname*`.rept` output file. A score-in-place calculation writes a *jobname*`.scor` report file instead, and no structural output for the Pose Viewer. These results are retained as Maestro properties in the Pose Viewer file (*jobname*`_pv.mae`) or ligand database file (*jobname*`_lib.mae`), and can be displayed in the Maestro Project Table.

3.4.3 Liaison

Liaison uses its own scripts for running calculations from the command line. You can use Maestro to write the scripts, or you can modify the template scripts found in the directory:

`$SCHRODINGER/impact-v`*version_number*`/samples/liaison`

See the Maestro online help or the *FirstDiscovery Command Reference Manual* for more information.

The `simulate_`*jobname* script ensures that the "free" and "bound" input files are run for each ligand/receptor pair and are named appropriately. To run a Liaison calculation using job files you have written from the Maestro interface, enter the following in a terminal window:

`./simulate_`*jobname*

Once the Liaison simulations are complete, you can use the `analyze_`*jobname* script to do either of the following:

- Mine data and fit or predict mined results to known binding energies for the specific ligand/receptor pairs
- Predict the binding energies of the test ligands, given known values of alpha, beta, and gamma (typically those calculated by a previous fit)

However, it is more convenient to run the very fast fit and predict calculations directly from the Maestro interface.

Distributed processing of Liaison simulation jobs is available from the GUI only. See Section 7.5.

3.4.4 QSite

QSite jobs have two input files: one for Impact, which runs the molecular mechanics part of the calculation, and one for Jaguar, which runs the quantum mechanical part. The Jaguar component can be run in parallel if multiple processors are available, either from the command line or from the GUI. The command line switches, optional and required, for running QSite calculations are described in Table 3.4.

The syntax for specifying the two input files is:

`$SCHRODINGER/impact -j` *jag_jobname*`.in -i` *jobname*`.inp`

It is important that *jag_jobname* differ from *jobname*, so that working files and directories for the two programs do not collide, so Maestro does this automatically. For a QSite job set up from a session in which you enter the job name *jobname*, Maestro names the Impact input file *jobname*`.inp` and the Jaguar input file *jobname*`.jaguar.in`.

Table 3.4. QSite Command Line Options.

Syntax Example	Description
`-j` *jag_jobname*`.in`	Required: specifies the Jaguar input file *jag_jobname*`.in`
`-p` *num*	Specifies the number of processors to use for Jaguar
`-i` *jobname*`.inp`	Required: specifies the Impact input file *jobname*`.inp`
`-o` *logfile*	Specifies the file name for standard output and standard error. If omitted, *jobname*`.log` is used.

3.4.5 Basic Impact

Basic Impact calculations can be started at the command prompt using the syntax shown below:

```
$SCHRODINGER/impact -i jobname.inp [-o logfile]
```

Basic Impact jobs require only the `-i` option to specify the input file. If you want the log file to be written to a different file name than *jobname*`.log`, you can designate a specific file name with `-o` *logfile*.

For distributed processing, the `run_jobs.pl` script can be used. (Distributed processing for Basic Impact calculations is not available from the GUI.)

3.4.6 Using Job Control Commands

Once your jobs are launched, you can monitor their progress using the Monitor panel in Maestro. The command `$SCHRODINGER/jobcontrol` can also be used. It has many options, but the two most useful options are:

```
$SCHRODINGER/jobcontrol -list
```

which will show the status of all your jobs, and:

```
$SCHRODINGER/jobcontrol -kill
```

to terminate any job and its subjobs, if any exist.

For a summary of `jobcontrol` options, use:

```
$SCHRODINGER/jobcontrol -h
```

For more information, see the *Maestro User Manual*. For an introduction to running and monitoring jobs in Maestro, see Section 2.10 on page 24.

3.5 Using Command-Line Utilities

Several FirstDiscovery support modules are command-line applications or utilities: structure preparation utilities and applications, structure conversion utilities, and Glide utilities. The command-line application `protprep` is located in the main Schrödinger directory. Utilities are located in the directory `$SCHRODINGER/utilities`. You may want to add this directory to your path so that they are easy to run by name from the command line. For usage summary information, use the `-h` (help) option:

```
$SCHRODINGER/protprep -h
```

```
$SCHRODINGER/utilities/utilityname -h
```

Chapter 4: Protein Preparation

4.1 Protein and Ligand Structure Preparation

Because the quality of results depends on reasonable starting structures, Schrödinger offers a comprehensive protein preparation facility designed to ensure chemical correctness and to optimize protein structures for use with FirstDiscovery. It is strongly recommended that you process protein structures with the preparation facility in order to achieve best results.

This chapter describes the preparation of protein-ligand complexes using the FirstDiscovery protein preparation facility. Most features of the facility are available from the ProteinPrep panel. Additional features are available in the command-line application `protprep`. The utilities `pprep` and `impref` are also available. Use of the command-line application and utilities is summarized in Section 4.11 on page 55.

4.2 The ProteinPrep Panel

The ProteinPrep panel is intended to help in preparing protein structures from PDB files for use in the FirstDiscovery applications Glide, Liaison, and QSite. A typical structure downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB) website (http://www.rcsb.org) will have no hydrogens and may have residues in unusual charge states. FirstDiscovery uses an all-atom force field (OPLS-AA) and requires correct charge states near the binding site. Protein preparation for FirstDiscovery is mainly a process of neutralizing appropriate amino acid sidechains, adding hydrogens, orienting hydroxyl hydrogens, and relieving steric clashes.

Protein preparation takes place in two parts. After ensuring chemical correctness, the preparation process neutralizes side chains that are not close to the binding cavity and do not participate in salt bridges. The refinement portion performs a restrained minimization of the cocrystallized complex, which reorients side-chain hydroxyl groups and alleviates potential steric clashes. The ProteinPrep panel allows you to specify which parts of the procedure to run: Preparation and refinement, Preparation only, or Refinement only.

If you are familiar with Maestro, or if you have followed Schrödinger's protein preparation procedure before, you might need to follow only the overview provided in Section 4.3. The steps are described in detail in later sections of this chapter. For a tutorial on protein preparation, see the *FirstDiscovery Quick Start Guide*.

4.3 Step-by-Step Overview

This section provides an overview of the protein preparation process. The procedure described assumes that the initial protein structure is in a PDB-format file, includes a cocrystallized ligand, and does not include explicit hydrogens. For best results, structures with missing residues near the active site should be repaired before protein preparation. After processing with Schrödinger's protein preparation facility, you will have files containing refined, hydrogenated structures of the ligand and the ligand-receptor complex. The prepared structures are suitable for use with any FirstDiscovery application. In most cases, not all of the steps outlined need to be performed. See the descriptions of each step to determine whether it is required.

1. Import a ligand/protein cocrystallized structure, typically from PDB, into Maestro.
2. Locate any waters you want to keep, then delete all others. Generally, all waters (except those coordinated to metals) are deleted, but waters that bridge between the ligand and the protein are sometimes retained.
3. Simplify multimeric complexes.
   - Determine whether the protein-ligand complex is a dimer or other multimer containing duplicate binding sites and duplicate chains that are redundant.
   - If the structure is a multimer with duplicate binding sites, remove redundant binding sites and the associated duplicate chains by picking and deleting molecules or chains in Maestro.
4. Adjust the protein, metal ions, and cofactors.
   - Fix any serious errors in the protein. Incomplete residues are the most common errors, but are relatively harmless if they are distant from the active site. Structures that are missing residues near the active site should be repaired.
   - Check the protein structure for metal ions and cofactors.
   - Set charges and correct atom types for any metal atoms, as needed.
   - Set bond orders and formal charges for any cofactors, as needed.
5. Adjust the ligand bond orders and formal charges.
6. Run protein preparation.
   - Open the ProteinPrep panel, mark the ligand, choose the desired Procedure, and click Start.
7. Review the prepared structures.
   - If problems arise during the preparation or refinement stages, review the log file, correct the problems, and rerun.
   - Examine the refined ligand/protein/water structure for correct formal charges

and protonation states resulting from Step 6 and make final adjustments as needed.

4.4 Importing the Protein Complex Structure

This step begins the protein preparation procedure.

For an introduction to the Maestro interface, see Chapter 2. For help with the ProteinPrep panel or any Maestro command or procedure, see the online help or the *Maestro User Manual*. To reverse your most recent Maestro action, you can click the Undo button on the toolbar:

To import a ligand-receptor protein complex structure into Maestro:

1. On the toolbar, click the Import structures button:

   The Import panel is displayed.
2. Select PDB format.
3. Enter the name of the file, or select the file in the Files list.
4. Click Import.
5. To display the Project Table, click the Show/Hide project table button on the toolbar:

The imported entry is highlighted in the Project Table and visible in the Workspace.

4.5 Deleting Unwanted Waters

Water molecules in the crystallographic complex are generally not used unless they are judged critical to the functioning of the protein–ligand interaction. When waters are used, they are later included in the protein as “structural” waters.

If you know there are no waters that are important to the protein-ligand interaction, skip to Section 4.5.3 to delete all waters.

4.5.1 Locating Structural Waters

To probe for structural waters in the protein structure:

1. Locate the ligand and determine its molecule number.

   The ligand is usually evident when PDB structures are imported into Maestro, because structures that are not standard amino acid residues are colored orange, while complete standard residues are gray. If the ligand is not evident:

   a. On the toolbar, choose Chain Name or Molecule Number from the Color all atoms by scheme button menu:

   Locate the ligand.

   b. From the Display menu, choose Atom Labels.

   The Atom Labels panel opens.

   c. Make sure the selected Mode is Add and the Label Atoms pick state is Atoms.

   d. In the Composition folder, make sure that Molecule Number is the only option selected.

   e. Click on an atom of the ligand.

   The ligand's molecule number is shown.

2. In the Commands text box, enter:

   `displayonlyatom fillres within` *dist* `mol.num` *molnumlig*

   where *molnumlig* is the molecule number determined in Step 1. This will display the ligand plus all atoms (including water oxygens) within the chosen distance, *dist* Å, of the ligand.

   The display of whole residues with any atoms within the chosen distance can also be accomplished using the Atom Selection dialog box (ASD). For more information, see the online help or the *Maestro User Manual*.

3. To visualize the resulting waters:

   a. From the Display menu, choose Molecular Representation.

   b. In the Atoms folder, choose the Representation option Ball & Stick.

   c. Click Select to open the ASD.

   d. In the Residue folder, select Residue Type.

e. Click the water residue type, HOH.

f. Click Add.

g. Click OK.

The water oxygens, assuming no hydrogens have been added, are displayed in ball and stick representation as balls.

Alternatively, you can enter:

```
repatom rep=ballnstick res. HOH
```

in the Commands text box.

4.5.2 Should Structural Waters Be Kept?

Keeping structural waters is likely to be more important for Liaison than for Glide. Deleting all waters to make a site more accessible might be desirable in a Glide project. For example, retaining the water under the flap in 1HPX would prevent the docking of the DuPont-Merck cyclic urea that displaces this water. In other cases, removing waters might enable you to find ligands capable of "replacing" the missing waters. An alternative approach for 1HPX would be to prepare versions of the protein with and without the flap water and to dock ligands against both. In our database screens for 1HPX, excellent rank orders were found for the known ligands, even though the flap water was removed.

4.5.3 Deleting All Water Molecules

If you decide to delete all waters, choose Waters from the Delete button menu on the toolbar. All water molecules are deleted. Skip to Section 4.6.

4.5.4 Deleting Distant Water Molecules

If you think you may want to keep one or more waters, begin by removing those that are farther than *dist* Å from the ligand.

This task can be performed using the Commands text box in the lowest part of the Maestro main window to enter a command such as:

```
delete res. HOH and beyond dist mol.n molnumlig
```

where *molnumlig* is the molecule number of the ligand. Alternatively, use the Delete button and the Atom Selection dialog box (ASD):

1. Choose Select from the Delete button menu on the toolbar.

   The Select to Delete ASD is opened.

2. In the Molecule folder, choose Molecule Number and enter the ligand's molecule number. Click Add.
3. Click the Proximity button in the lower section of the ASD. In the Proximity dialog box:
   a. Select Beyond, enter a distance in the text box, and select Angstroms.
   b. Under Fill, select Residues and Exclude source. This keeps the ligand itself from being deleted.
   c. Click OK to exit the Proximity dialog box.
4. In the Select to Delete ASD, click the Residue tab, choose Residue Type, and select HOH. Click Intersect.

   The ASL box will contain an expression similar to:

   ```
   (not (mol.num 2) and fillres beyond 5 (mol.num 2 ) ) AND
   ((res.ptype "HOH "))
   ```

   and most of the water oxygens are marked in the Workspace.
5. Click OK to delete the selected water molecules.

4.5.5 Delete Remaining Unwanted Waters

After deleting water molecules beyond distance *dist* Å from the ligand, examine the Workspace and delete any remaining water molecules you do not want to keep:

1. On the toolbar, choose Molecules from the Delete button menu:

2. Click on an oxygen to delete that water molecule.

   When you have removed all but the desired waters, continue with Section 4.6.

4.6 Simplifying the Protein Complex

4.6.1 Determining Whether the Complex Is a Multimer

To determine whether the ligand-receptor complex is a multimer, compare the chains that appear in the sequence viewer. If there are two or more chains with identical sequences, the complex may be a multimer. If this is the case, there may be duplicate copies of the binding site of interest, with duplicate chains forming the duplicate binding sites.

If the binding interaction of interest takes place within a single subunit, you should retain only the one ligand-receptor subunit to prepare for Glide, Liaison, or QSite. However, if two identical chains are both required to form the active site, neither should be deleted. To see whether two duplicate chains are involved with the active site, undisplay the protein's amino acid residues:

1. On the toolbar, choose Protein Backbone from the Undisplay button menu:

2. Repeat the process and choose Protein Side Chains.

Ligands, cofactors, metal ions, and water-molecule oxygens remain visible. If two or more identical ligands or ligand/cofactor groups are present, then the complex is most likely a multimer, and the redundant groups and the duplicate chains associated with them can be deleted.

4.6.2 Retaining Needed Subunits

If the protein complex structure is a multimer with duplicate binding sites, it can be truncated by deleting all but a single ligand binding site and the associated receptor subunit(s). If you choose not to truncate the structure, skip to Section 4.7 on page 45.

To remove redundant subunits or receptor sites of a multimer:

1. Delete all but one ligand or ligand/cofactor pairing:
   a. On the toolbar, choose Molecules from the Delete button menu:

   b. Click on any atom in a molecule to delete that molecule.
2. Display the ligand or ligand/cofactor pair in CPK:
   a. On the toolbar, choose Molecules from the Draw atoms in CPK button menu:

   b. Click on an atom in the ligand that was not deleted to display it in CPK.
   c. If there is a cofactor, click on an atom in that molecule as well.
   d. Click the toolbar button a second time to leave the Draw atoms in CPK pick state.

3. Redisplay the protein backbone:

   On the toolbar, choose Protein Backbone from the Also display button menu:

   Making just the backbone visible will provide enough information without unduly cluttering the Workspace.

4. Assign coloring by Chain Name:

   On the toolbar, choose Chain Name from the Color all atoms by scheme button menu:

5. Delete duplicate protein chains:

   a. On the toolbar, choose Chains from the Delete button menu.

   b. Click on a backbone atom in each protein chain you want to delete.

6. Delete duplicate ligands and cofactors:

   a. On the toolbar, choose Molecules from the Delete button menu.

   b. Click on an atom in each ligand or cofactor to be deleted.

   Note: If you make a mistake, you can undo your last action by clicking the Undo button on the toolbar:

7. When finished, redisplay the rest of the protein:

   On the toolbar, choose All from the Display only button menu:

8. Put all atoms, including the ligand and any cofactors, back into wire-frame:

   On the toolbar, double-click the Draw bonds in wire button:

4.7 Adjusting the Protein, Metal Ions, and Cofactors

4.7.1 Proteins That Already Include Hydrogen Atoms

If the protein already includes hydrogen atoms, you will need to decide how to proceed. If all hydrogens are present, you could use the structure as is and omit running the protein preparation procedure. This approach is not recommended unless you are absolutely satisfied that the structure is properly prepared and contains no untenable steric clashes. Otherwise, follow the steps provided below to continue the preparation procedure.

4.7.2 Checking the Protein Structure for Metal Ions and Cofactors

1. Ensure that the protein, with metals and cofactors, is included in the Workspace.
2. To help find any metal ions or cofactors, recolor the atoms by element:

   On the toolbar, choose Element from the Color all atoms by scheme button menu:

   All atoms in the Workspace are now colored by element.
3. Examine the protein structure to determine how to continue.
   - If the protein contains neither metal ions nor cofactors, proceed to Section 4.8.
   - If the protein contains metal ions but no cofactors, continue with Section 4.7.3 and check metal ion properties before proceeding to Section 4.8.
   - If the protein contains cofactors but no metal ions, continue with Sections 4.7.4 and 4.7.5 and check cofactor properties before proceeding to Section 4.8.
   - If the protein contains both metal ions and cofactors, do all procedures: check metal ion properties as described in Section 4.7.3, check cofactor properties as described in Sections 4.7.4 and 4.7.5, then proceed to Section 4.8.

4.7.3 Adjusting Metal Ions

Metal ions in the protein complex cannot have covalent bonds to protein atoms. Preparation jobs run from the ProteinPrep panel now automatically delete protein-metal bonds. Therefore there is usually no need to explicitly delete the bonds before checking and adjusting element names and formal charges.

To manually delete bonds between metals and protein atoms (where necessary):

1. On the toolbar, click-hold the Delete button, selecting Bonds from the button menu.
2. Click on the bonds to be deleted.

The MacroModel atom types for metal ions are sometimes incorrectly translated into dummy atom types (Du, Z0, or 00) when metal-protein bonds are specified in the input structure. Furthermore, isolated metal ions may erroneously be assigned general atom types (GA, GB, GC, etc.). The ProteinPrep procedure cannot treat structure files containing these atom types; they should be corrected as described in this section.

To display element labels and formal charges:

1. Open the Build panel by clicking the Show/Hide build panel button:

2. In the Build toolbar, click the Label button:

   All metal ions (and other heteroatoms) are labeled with their element symbol and formal charge.

3. Check any metal ions to make sure they are correct. If they are, the next step in the process is Section 4.7.4. If not, you can correct them.

To correct metal ion atom types:

1. In the Build panel, click the Atom Properties tab and select Atom Type (MacroModel) from the Property option menu.
2. Find the correct atom type for the metal ion.

   The atom type for metal ions includes both element name and formal charge. Atom type numbers are in parentheses.

3. Click in the list to select the correct atom type.
4. Click on the metal ion to be changed.

4.7.4 Displaying the Cofactor

Cofactors are included as part of the protein, but because they are not standard residues it is sometimes necessary to use Maestro's structure-editing capabilities to ensure that multiple bonds and formal charges are assigned correctly.

To display only the cofactor:

1. On the toolbar, choose Select from the Display only button menu:

   The Atom Selection dialog box (ASD) is displayed.

2. In the Residue folder, choose Residue Type.
3. Click the residue type of the cofactor, which will be near the end of the list.

   The cofactor is highlighted.

4. Click Add, then click OK.

The cofactor is displayed. Because the cofactor was chosen by residue type and not molecule number, this method works even if the cofactor is covalently bonded to another residue.

4.7.5 Adjusting Cofactor Atom and Bond Properties

To set or change cofactor bond orders:

1. If the Build panel is not displayed, click Show/Hide build panel on the main toolbar:

2. On the Build panel toolbar, click the Decrement bond order or Increment bond order button, as appropriate:

3. Click on bonds as necessary to set the bond order.

To set or correct the formal charge on any cofactor atoms:

1. On the main toolbar, choose Formal Charge from the Label atoms button menu.

2. If the Build panel is not displayed, click Show/Hide build panel on the main toolbar:

3. On the Build panel toolbar, click on the Increment formal charge or Decrement formal charge button, as appropriate:

4. Click on an atom whose formal charge must be increased or decreased. Repeat as necessary. The atom labels will show the current formal charge.

To correct the atom type of any mistyped atoms:

1. If the Build panel is not displayed, click Show/Hide build panel on the main toolbar:

2. In the Atom Properties folder of the Build panel, choose Atom Type (MacroModel) from the Property option menu.
3. Find the correct atom type for the mistyped atom, and click it in the list.
4. Click on the atom to be changed.
5. If the cofactor contains any metal ions, bonds between the metal and cofactor can be removed as in Section 4.7.3.

For more information about structure editing in Maestro, click Help or see the *Maestro User Manual*.

4.8 Adjusting the Ligand

4.8.1 Adjusting Ligand Atom and Bond Properties

If you have not already colored by element, do so now:

Choose Element from the Color all atoms by scheme button menu:

To set or change ligand bond orders:

1. If the Build panel is not displayed, click Show/Hide build panel on the main toolbar:

2. On the Build panel toolbar, click the Decrement bond order or Increment bond order button, as appropriate:

3. Click on bonds as necessary to set the bond order.

To set or change formal charges on any ligand atoms:

1. On the main toolbar, choose Formal Charge from the Label atoms button menu:

2. If the Build panel is not displayed, click Show/Hide build panel on the main toolbar:

3. On the Build panel toolbar, click on the Increment formal charge or Decrement formal charge button, as appropriate:

4. Click on an atom whose formal charge must be increased or decreased. Repeat as necessary. The atom labels show the current formal charge.

To correct the atom type of any mistyped atoms:

1. On the main toolbar, choose Atom Type (MacroModel) from the Label atoms button menu:

2. If the Build panel is not displayed, click Show/Hide build panel on the main toolbar:

3. In the Atom Properties folder of the Build panel, choose Atom Type (MacroModel) from the Property option menu.
4. Find the correct atom type for the mistyped atom, and click it in the list.
5. Click on the atom to be changed.

4.8.2 Manually Deleting Explicit Ligand-Metal Bonds

If the complex structure contains any bonds from the ligand or a cofactor to a protein metal, they must be deleted. Glide, and the OPLS-AA force field it uses, models such interactions as a van der Waals plus electrostatic interaction. Glide cannot handle normal covalent bonds to the ligand, such as might be found in an acyl enzyme.

Maestro will delete all bonds to metal atoms before the protein preparation job begins. However, if there are metal-ligand bonds, they will interfere with the identification of the ligand molecule. To avoid this, delete such bonds manually.

1. To check for ligand-metal bonds, you can use Maestro's Display/Undisplay facility.
   a. On the toolbar, choose Protein Backbone from the Undisplay button menu:

   b. Repeat the process, choosing Protein Side Chains.
2. If any metal-ligand bonds exist, delete them:
   a. Choose Bonds from the Delete button menu on the toolbar.
   b. Click on the bonds to be deleted.
3. Redisplay the complete protein by choosing All from the Display Only button menu on the toolbar.

4.8.3 Checking for Other Protein-Ligand Bonds

If any covalent bonds exist between the protein and ligand, they will preclude treatment in Glide or Liaison. Bonds between the components of the protein/metals/cofactor structure were deleted in Section 4.7.3.

4.9 Running ProteinPrep on the Structures

From this point on, all structural manipulations are done by the ProteinPrep panel, shown in Figure 4.1, and its related scripts. Before you open this panel, ensure that the protein and ligand are in the Workspace.

To open the ProteinPrep panel, select ProteinPrep from the Applications menu on the main menu bar.

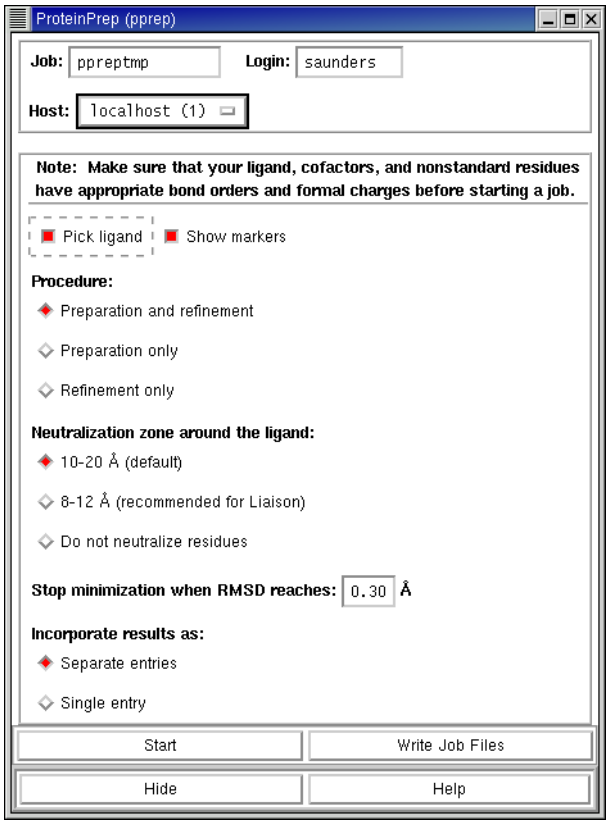


Figure 4.1. The** ProteinPrep **panel.

4.9.1 Entering Job Settings

To enter job settings:

1. Type a name for the job in the Job text box.
2. Choose a host for the job from the Host option menu.

   If you use a different login on the remote host you have selected, enter that login in the Login text box.

4.9.2 Defining the Ligand

Before launching a protein preparation job, you must choose a molecule in the Workspace that will be treated as the ligand. In the ProteinPrep panel, choose Pick, and then select the ligand by clicking on it in the Workspace. When Show markers is chosen, the ligand will be highlighted with a blue-green marker. The rest of the Workspace is then treated as the protein.

4.9.3 Choosing a Procedure

The ProteinPrep panel facilitates three types of jobs: Preparation only, Refinement only, and Preparation and refinement.

The Preparation component neutralizes residues that are beyond a set distance from the ligand. The Preparation process also detects some conflicts in hydrogen bonding. It corrects them when possible, either by exchanging carbonyl and hydroxyl oxygens in a neutralized carboxylic acid group, or by creating the alternate (HIE) tautomer of a histidine side chain.

The Refinement component uses Impact to run a series of restrained, partial minimizations on the combined, hydrogenated structure. Minimizations continue until the average RMS deviation of the non-hydrogen atoms reaches the specified limit (0.3 Å by default).

The first step in the sequence of restrained minimizations reorients side-chain hydroxyls in serine, threonine, and tyrosine residues, and side-chain sulfhydryls of cysteine residues. This is accomplished by tightly tethering non-hydrogen atoms with a force constant of 10 kcal/mol·Å^2 and by minimizing the hydrogens with torsion interactions turned off.

Each restrained minimization employs a limited number of minimization steps and is not intended to minimize the system completely. Subsequent steps restore the torsion potential and use progressively weaker restraints on the non-hydrogen atoms (hydrogen atoms are always free). The force constants employed are 3, 1, 0.3, and 0.1 kcal/mol·Å^2.

Preparation and refinement, the default, runs both components. This is the recommended mode if you have not yet run any preparation jobs on the protein. Separate Preparation only and Refinement only jobs can be run if you encountered a problem in the combined Preparation and refinement job. Subsequent Refinement only jobs can be performed after a Preparation and refinement job if water molecules need to be reoriented or if other structural adjustments need to be made.

4.9.4 Other Options

1. Neutralization zone around the ligand: 10-20 Å

   You can also choose a shorter distance (8-12 Å) or choose Do not neutralize residues.

2. Stop minimization when RMSD reaches: 0.30 Å

   This is the default value. It allows the refinement portion of the job to halt when the average RMSD of the heavy atoms reaches 0.30 Å.

3. Incorporate results as: Separate entries.

   This setting determines how the output structure is included in the Maestro Project Table upon job completion. Separate entries returns the protein and ligand as separate entries, whereas Single entry does not. Single entry may be more convenient if you want to run calculations on the complex (with the existing ligand) immediately.

4.9.5 Launching the ProteinPrep Job

To launch a ProteinPrep job, click Start. The monitor panel is displayed, and the results of the job are shown. If you decide to run preparation and refinement separately, you will need to run a Preparation only job and refine the results with a subsequent Refinement only job. If you want to run the job from the command line, click Write Files. See Section 4.11 on page 55 for information on command line options.

4.9.6 Output Job Files

Running Preparation and refinement produces the same files as running Preparation only followed by Refinement only. The following structure files are produced, where *struct* is the name of the complex:

*struct*_lig.mae	The input ligand structure file
*struct*_lig_prep.mae	The post-preparation ligand structure file

*struct*_lig_ref.mae	The post-refinement ligand structure file
*struct*_prot.mae	The input receptor structure file
*struct*_prot_prep.mae	The post-preparation receptor structure file
*struct*_prot_ref.mae	The post-refinement receptor and ligand structures
struct.log	The log file for the complete preparation and refinement job

4.10 Checking the Output Structures

Finally, after both the preparation and refinement components have successfully run, you should check the completed ligand and protein structures.

4.10.1 Checking the Orientation of Water Molecules

Perform this step only if you identified and kept some structural waters in Section 4.5. Reorienting the hydrogens is not strictly necessary, as their orientation should have been changed during refinement in Section 4.9, but it is useful to check that the orientation is correct.

If the orientation is incorrect, reorient the molecules by using the following procedure:

1. On the toolbar, choose Global/Local from the Local transformation button menu:

   The Advanced Transformations panel is displayed.
2. Under Atoms For Transformation, use the picking controls to select the entire water molecule you want to reorient.
3. Under A Center For Transformation, use the picking controls to select the oxygen atom of the water molecule.
4. Under Rotation/Translation Scope, select Local.
5. Use the middle mouse button to change the orientation of the water hydrogens.
6. Hide the Advanced Transformations panel. Transformations should now be global again.

When you have corrected the orientation of the retained water molecules, run a Refinement only job on the adjusted protein-ligand complex as described in Section 4.9.

4.10.2 Resolving H-Bonding Conflicts

One or more residues may need to be modified to resolve an acceptor-acceptor or donor-donor clash. If residues need to be modified, follow these steps:

1. Place the refined protein-ligand complex in the Workspace.
2. Examine the interaction between the ligand and the protein (and/or the cofactor).
3. Use your judgment and chemical intuition to determine which protonation state and tautomeric form the residues in question should have.
4. Use the structure-editing capabilities in Maestro to resolve the conflict.

Some of these clashes are recognized by the preparation process but cannot be resolved by it. The preparation process may have no control over other clashes. An example of the latter typically occurs in an aspartyl protease such as HIV, where both active-site aspartates are close to one or more atoms of a properly docked ligand. Because these contact distances fall within any reasonable cavity radius, the carboxylates are not subject to being neutralized and will both be represented as negatively charged by the preparation process. However, when the ligand interacts with the aspartates via a hydroxyl group or similar neutral functionality, one of the aspartates is typically modeled as neutral.

4.11 Command-Line Protein Preparation

To run protein preparation from the command line:

1. If you do not yet have receptor and ligand structure files for the structures in the Workspace, use the Write Job Files button in the ProteinPrep panel to write the structure files.
2. Use the `protprep` command-line application or the `pprep` and/or `impref` command-line utilities to run specific procedures. These commands and their options are summarized below.

4.11.1 Usage Summary for `protprep`

The `$SCHRODINGER/protprep` application has command-line options corresponding to features of the Maestro ProteinPrep panel. The command `protprep -h` displays the usage summary that appears in this section.

Syntax:

`$SCHRODINGER/protprep [`*options*`]` *input-file*

input-file is the file containing the protein to be prepared or refined. This file must be in Maestro format. When doing a refinement only job (`-mode refine`) this file can contain a protein-ligand complex.

Options:

General Options:

`-j` *jobname*	Override the default job name derived from *input-file*. This allows you to choose an output job name that is different from the *input-file* name.
`-l` *ligand-file*	Specify a file containing a ligand in the protein's active site. This file must be in Maestro format.
`-m` *mode* `-mode` *mode*	Select mode, where *mode* is one of the following: `prep` Preparation only. `refine` Refinement only. `both` Preparation and refinement (default).
`-debug`	Print verbose (debugging) output.
`-HOST` *host*	Run the job on a remote host.
`-LOCAL`	Run the job in the current directory, rather than in a temporary scratch directory.
`-WAIT`	Keep job in foreground. Do not return until the job completes. (The default is to run `protprep` in the background.)
`-NICE`	Run the job at reduced priority.
`-HELP | -h`	Print usage message and exit.

Preparation Stage Options:

`-min-recep-only`	Minimize total charge of receptor only.
`-skip-sidechain-corr`	Skip correction of conflicting side-chain forms.
`-cavity-8-12`	Set cavity distance range to 8-12 Å. Suitable for Liaison jobs.
`-salt-bridge-dist`	Leave residue pairs forming salt bridges within *dist* Å ionized; default is 3.5 Å.
`-ionization-range`	Leave residues within *dist* Å of ligand ionized.
`-hbond-dist`	Set H-bonding distance; default is 3.45 Å.

Refinement Stage Options:

`-r` *rmsd*	Maximum RMSD allowed for refinement; default 0.3.
`-keep`	Keep intermediate structure files.
`-separate`	Write out refined protein and ligand structures separately, rather than in one combined structure.

4.11.2 Usage Summary for `pprep`

Note: `pprep` is the driver for the preparation stage, and is invoked by `protprep`. There is little need to run `pprep` directly.

Purpose: To adjust protonation states of a receptor in a Maestro format file.

Syntax:

```
$SCHRODINGER/utilities/pprep [options] proteinfile.mae
```

Options:

`-i` *idis*	Leave residue pairs forming salt bridges within *idis* ionized; default is 3.5 Å.
`-l` *ligfile*	Read ligand mae file *ligfile*.
`-n` *outfile*	Specify non-default (*lig*`R.mae`) name for output file with neutralized residues.
`-p`	Print verbose output.
`-r`	Minimize total charge of the receptor only.
`-t`	Skip correction of conflicting side-chain forms.
`-w` *wdis*	Leave residues within *wdis* of ligand ionized.
`-H` *hbonddist*	Set H-bonding distance; default 3.45 Å.
`-L`	Set cavity distance range to 8-12 Å.
`-v`	Print version number and exit.
`-h`	Print usage message and exit.

4.11.3 Usage Summary for `impref`

Note: `impref` is the driver for the refinement stage, and is invoked by `protprep`. There is little need to run `impref` directly.

Purpose: To use Impact for restrained optimizations of a ligand-receptor complex.

Syntax:

```
$SCHRODINGER/utilities/impref [options] input.mae
```

Options:

`-k`	Keep Impact minimization `*.inp`, `*.log`, and `*.mae` files.
`-l` *ligfile*	Read ligand from file *ligfile*, instead of *input*`.mae`. If this option is used, *input*`.mae` must be the protein structure alone. If this option is not used, *input*`.mae` must be the protein/ligand complex.
`-r` *rmsd*	Specify maximum RMSD allowed; default is 0.3.
`-s`	Write out protein and ligand separately. Requires `-l` *ligfile*.
`-op` *file*	Output protein or complex file. Default is *input*`_ref.mae`.
`-ol` *file*	Output ligand file (when `-s` and `-l` used.) Default is *ligfile*`_ref.mae`.
`-v`	Print version number and exit.
`-h`	Print usage message and exit.

Chapter 5: Ligand Preparation

5.1 Ligand Preparation Checklist

Chapter 4 discussed the preparation of receptor and ligand-receptor structure files for use in FirstDiscovery. Candidate ligand structures must also have certain characteristics for FirstDiscovery applications. Some of these conditions can be met by using Maestro features or command-line utilities to alter the ligand structure file.

To be submitted to Glide and other applications, ligand structures:

1. Must be three-dimensional (3D).
2. Must each consist of a single molecule that has no covalent bonds to the receptor, with no accompanying fragments, such as counter-ions and solvent molecules.
3. Must be in a Maestro-format file. Maestro transparently converts SD, MacroModel, and PDB formats to Maestro format during structure import. Maestro also transparently converts Mol2 during import. However, FirstDiscovery has no direct Mol2 support, so make sure your structures are in Maestro, SD, or PDB format before beginning FirstDiscovery jobs.

   (Structure file format conversion can be done from the command line using utilities such as `pdbconvert`, `sdconvert`, and `maemmod`. See the *Maestro User Manual*.)
4. Must have all their hydrogens (filled valences). These can be added in Maestro by using either the Add hydrogens toolbar button:

   or the Hydrogen Treatment panel (select Hydrogen Treatment from the Edit menu).

   Hydrogen atoms can also be added (or removed) using the command-line tool `applyhtreat`, which is described in the *Maestro User Manual* and the *LigPrep User Manual*.

5.2 LigPrep

The Schrödinger ligand preparation product LigPrep is designed to prepare high quality, all-atom 3D structures for large numbers of drug-like molecules, starting with 2D or 3D structures in SD or Maestro format. LigPrep can be run from Maestro or from the command line.

LigPrep is licensed separately from other Schrödinger products. The MacroModel commands `premin` and `bmin` require LigPrep licenses when run in a LigPrep context, and are limited to a restricted set of commands when run using a LigPrep license. For more information about obtaining LigPrep, contact help@schrodinger.com.

The simplest use of LigPrep produces a single low-energy 3D structure with correct chiralities for each successfully processed input structure. LigPrep can also produce a number of structures from each input structure with various ionization states, tautomers, stereochemistries, and ring conformations, and eliminate molecules using various criteria including molecular weight or specified numbers and types of functional groups present.

5.2.1 The LigPrep Process

The LigPrep process consists of a series of steps that perform conversions, apply corrections to the structures, generate variations on the structures, eliminate unwanted structures, and optimize the structures. Many of the steps are optional, and are controlled by selecting options in the LigPrep panel or specifying command-line options. The steps are outlined below. Each step is performed by the script or program listed in the step.

1. Convert structure format.

   If the input structure file is in SD format it is converted to Maestro format by `sdconvert`. Parities specified in the SD file are converted into chiralities, which are stored as properties in the Maestro file.

2. Select structures.

   A subset of the input structures can be selected for processing. The selection is done by `maesubset` for Maestro input files and by `sdconvert` for SD input files.

3. Add hydrogen atoms.

   Structures that have implicit hydrogen atoms may need to have hydrogen atoms added before the 3D structures can be minimized. Hydrogen atoms are added in a manner that is consistent with a particular force field. This step is performed by `applyhtreat`, which is the program used by the Hydrogen Treatment panel in Maestro.

4. Remove unwanted molecules.

   If structures have additional molecules included, such as counter ions in salts and water molecules, these may need to be removed. The `desalter` removes all but the molecule containing the most atoms from each structure.

5. Neutralize charged groups.

Charged groups must be neutralized before ionization states can be generated. Neutral molecules are also required by various applications, such as QikProp. The neutralization is performed by `neutralizer`, which adds or removes hydrogen atoms.

6. Generate ionization states.

   For some applications it is important that all species that exist in a given pH range are available. In this step, the `ionizer` generates various ionization states for each structure. This step should be preceded by a neutralization step.

7. Generate tautomers.

   As with ionization, the significantly populated tautomers may be important for some types of calculations, such as docking with Glide. The `tautomerizer` generates various tautomers for each structure.

8. Filter structures.

   In this step, structures that match specified conditions can be removed. The condition can be on a property, such as Molecular weight > 1000, or on the structure, such as the presence or absence of a specific functional group. This step is performed by `ligparse`.

9. Generate alternative chiralities.

   2D structures do not always have complete chirality information, and it can be useful to vary the chiralities of the atoms to find all the low-energy structures or to provide a range of possible structures for investigation. This step identifies additional chiral atoms in the structures and generates additional structures with the same molecular formula but different chiral properties. The step is performed by the `stereoizer`.

10. Generate low-energy ring conformations.

    When ring conformation information is not available, it is important to generate a range of conformers so that the low-energy structures can be located. Ring confirmations are generated for each structure by `ring_conf`.

11. Remove problematic structures.

    Structures that could cause subsequent processing failures either in the energy minimization of the structures or in other applications are removed by `premin`.

12. Optimize the geometries.

    The geometries of the generated structures are optimized using a restricted version of the MacroModel computational program, `bmin`, or a short conformational search is performed to relax the structure into 3 dimensions while strongly encouraging chiral centers to adopt the proper chiralities (if the structure is highly strained).

13. Convert output file.

    If output in SD format was requested, `sdconvert` is run to perform the conversion.

5.2.2 The LigPrep Panel

The LigPrep panel allows you to set up LigPrep jobs in Maestro. Choose LigPrep from the Applications menu to open the panel. For details of panel options and operation, see the *LigPrep User Manual*.

Below are notes on panel options that produce more than one output structure per input structure.

The default options in the LigPrep panel run the `desalter`, add hydrogens, and minimize the ligand structure (performing a 2D-3D conversion, if necessary).

The `stereoizer` can generate two stereoisomers per chiral center in the ligand, up to a specified maximum. There are three Stereoisomers options:

The first two options, Retain specified chiralities (the default) and Determine chiralities from 3D structure, generate both isomers only at chiral centers where chirality is unspecified or indeterminate; centers with known chirality retain that chirality.

The difference is that Retain specified chiralities takes its chirality data from the input file (SD or Maestro), while Determine chiralities from 3D structure ignores input file chiralities and takes chirality information from the 3D geometry.

Generate all combinations produces the maximum number of structures, up to the maximum, which by default is 32 stereoisomers, but can be changed using Generate stereoisomers (maximum) *max* per ligand.

The `ionizer` (following the `neutralizer`) can generate all the ligand protonation states that would be found in the specified pH range. The Ionization options are:

Retain original state

Neutralize (best for QikProp)

Generate possible states at target pH *target* +/- *range.* This is the default, and can generate several different output structures for each input structure. The default pH *target* is 7.0 with a +/- *range* of 2.0, so the default pH range is 5.0-9.0. Both the target and range settings can be changed.

Generate stereoisomers (maximum): *max* per ligand.

Generate low energy ring conformations: *rings* per ligand.

Desalt is selected by default.

Generate tautomers is selected by default. The `tautomerizer` generates up to 8 tautomers per ligand, selecting the most likely tautomers if more than 8 are possible. If you are comfortable that the input structures are already in the correct tautomeric form for docking to a particular target, then the tautomerizer should be turned off by deselecting Generate tautomers.

5.3 The Ionization State Expander (`ionizer`)

While LigPrep as a whole requires additional licenses, one LigPrep tool, the `ionizer`, is included with FirstDiscovery. This section provides an introduction and usage summary for the `ionizer` as a service in FirstDiscovery.

The `ionizer` generates ionization states of ligands to match the pH range and other conditions you specify. The resulting ligands can be used as input to programs such as Glide. Starting with a Maestro-format input file of neutral molecular structures (for example, from a database), the `ionizer` produces a Maestro-format output file that has expanded to include multiple ionization states of each molecule, allowing Glide to select among them.

The `ionizer` requires the installation of a module called `services`. When you run the `INSTALL` script to install Schrödinger software, be sure to select the `services` product, which contains the `ionizer` software. For more information on installing FirstDiscovery and other Schrödinger software, see the *Schrödinger Product Installation Guide*.

The `ionizer` must be run from the command line as follows:

```
$SCHRODINGER/utilities/ionizer [options]
```

The options are listed in Table 5.1.

Table 5.1. Summary of `ionizer` *Options.*

Option	Description
`-h|-help`	Show this usage summary message.
`-doc`	Show more detailed usage message.
`-v|-ver|-version`	Show program version information.
`-j|-job|-jobname` *jobname*	Base name of job. No default (must be specified unless all essential files are specified).
`-i|-in|-infile` *infile*	Default is *jobname*`.mae`.
`-o|-out|-outfile` *outfile*	Default is *jobname*`-ion.mae`.
`-b|-bad|-badfile` *badfile*	Default is *jobname*`-ion-bad.mae`.

Table 5.1. Summary of `ionizer` *Options. (Continued)*

Option	Description
`-l|-log|-logfile` *logfile*	Default is *jobname*`.log`; use `-l` to log to screen.
`-ph` *value*	Effective pH of active site (default 7.0).
`-pht|-phthresh` *maxdiff*	pH difference threshold (default 2.0). For pH-based ion state rejections, where *maxdiff* is the difference limit on de/protonated |pKa-pH|.
`-pkt|-pkthresh` *maxdiff*	Strong/weak pK threshold (no default). Overrides pH-based rejection mode; reject on pKa values only (no pH), where *maxdiff* is the limit on de/protonated pKa differences.
`-mi|-maxions` *count*	Maximum number of ionizations (default 4).
`-mq|-maxabstotq` *charge*	Maximum absolute total charge (default 2).
`-mg|-maxgroups` *count*	Maximum number of ion groups to handle (default 15).
`-mo|-maxoutcts` *count*	Maximum number of output structures per input structure (default 512).
`-sm|-showmatches`	Show substructure pattern matches.
`-sf|-showfinal`	Show final ionization candidate list.
`-ll|-loglevel` *level*	Expansion report log level: Use `0` for quietest (default). Use `1` to log state generations. Use `2` to log ion fragment fusions too.
`-ss|-showskips`	Show skipped state generations. Augments log level 1 and up. Log levels > 1 give skip reasons.
`-kp|-keep_props`	Retain all properties in output CTs. Absent this option, connectivity-dependent properties are cleared
`-strict`	Terminate run if any input CT is bad. Unsets default fault-tolerant mode. Bad structure file option is ignored.
`-s|-spec|-specfile` *specfile*	Use nonstandard patterns spec file.
`-rw|-retitle_with` *prefix*	Add ion state number onto structure titles.

For a more detailed usage summary, use the command

```
$SCHRODINGER/utilities/ionizer -doc
```

For complete documentation on the `ionizer`, see the `README` file:

```
$SCHRODINGER/services-v version/doc/README.ionizer
```

Chapter 6: Glide

This chapter contains:

- A brief introduction to the Glide (Grid-based Ligand Docking with Energetics) program, its scientific methods and computational procedures.
- A detailed description of the Glide panel in Maestro and each of its folders, including instructions for using Glide constraints, Glide similarity, and extra-precision Glide docking (Glide XP).
- A description of the Glide Pose Viewer panel.
- A usage summary for Glide utilities, including `para_glide`, which permits distributed processing for docking large sets of ligands.

For more information on the technical aspects of Glide and on its performance and results, see the *FirstDiscovery Technical Notes*. For a tutorial on using Glide, see the *FirstDiscovery Quick Start Guide*. Much of the information in this chapter is available in the Maestro online help.

6.1 Introduction to Glide

Glide uses a hierarchical series of filters to search for possible locations of the ligand in the active-site region of the receptor. The shape and properties of the receptor are represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand poses. A *pose* is a complete specification of the ligand—its position and orientation.

Conformational flexibility is handled in Glide via an extensive conformational search, augmented by a heuristic screen that rapidly eliminates conformations deemed unsuitable for binding to a receptor, such as conformations that have long-range internal hydrogen bonds. As illustrated in Figure 6.1 on page 66, each ligand is divided into a *core* region and some number of *rotamer groups*. Each rotamer group is attached to the core by a rotatable bond, but does not contain additional rotatable bonds. The core is what remains when each terminus of the ligand is severed at the "last" rotatable bond. Carbon and nitrogen end groups terminated with hydrogen (—CH_3, —NH_2, —NH_3^+) are not considered rotatable because their conformational variation is of little significance. In Figure 6.1, the four central torsions are part of the core, and the methyl groups are not considered rotatable.

During conformation generation, each core region is represented by a set of core conformations, the number of which depends on the number of rotatable bonds, conformationally labile 5– and 6–membered rings, and asymmetric pyramidal trigonal nitrogen centers in the core. This set typically contains fewer than 500 core conformations, even for quite

large and flexible ligands, and far fewer for more rigid ligands. Every rotamer state for each rotamer group attached to the core is enumerated. The core plus all possible rotamer-group conformations is docked as a single object in Glide. Glide can also dock sets of pre-computed conformations. However, Glide offers its greatest value when flexible docking is used to generate conformations internally.

For each core conformation (or for rigid docking, each ligand), an exhaustive search of possible locations and orientations is performed over the active site of the protein. The search begins with the selection of "site points" on an equally spaced 2 Å grid that permeates the active-site region (Stage 1 in Figure 6.2 on page 67). To make this selection, pre-computed distances from the site point to the receptor surface, evaluated at a series of pre-specified directions and binned in 1 Å ranges, are compared to binned distances from the ligand center (the midpoint of the two most widely separated atoms) to the ligand surface. Glide positions the ligand center at the site point if there is a good enough match, but skips over the site point if there is not.

The second stage of the hierarchy begins by examining the placement of atoms that lie within a specified distance of the line drawn between the most widely separated atoms (the *ligand diameter*). This is done for a pre-specified selection of possible orientations of the ligand diameter (Step 2a). If there are too many steric clashes with the receptor, the orientation is skipped. Next (Step 2b), rotation about the ligand diameter is considered, and the interactions of a subset consisting of all atoms capable of making hydrogen bonds or ligand-metal interactions with the receptor are scored (*subset test*). If this score is good enough, all interactions with the receptor are scored (Step 2c). The scoring in these three tests is carried out using Schrödinger's discretized version of the ChemScore empirical

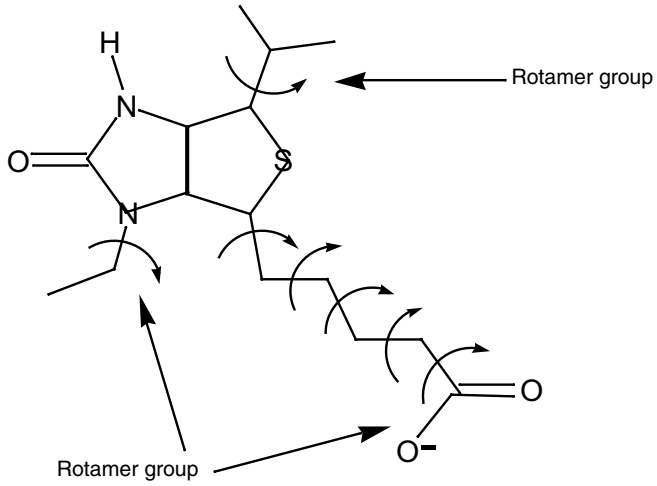


Figure 6.1. Definition of core and rotamer groups.

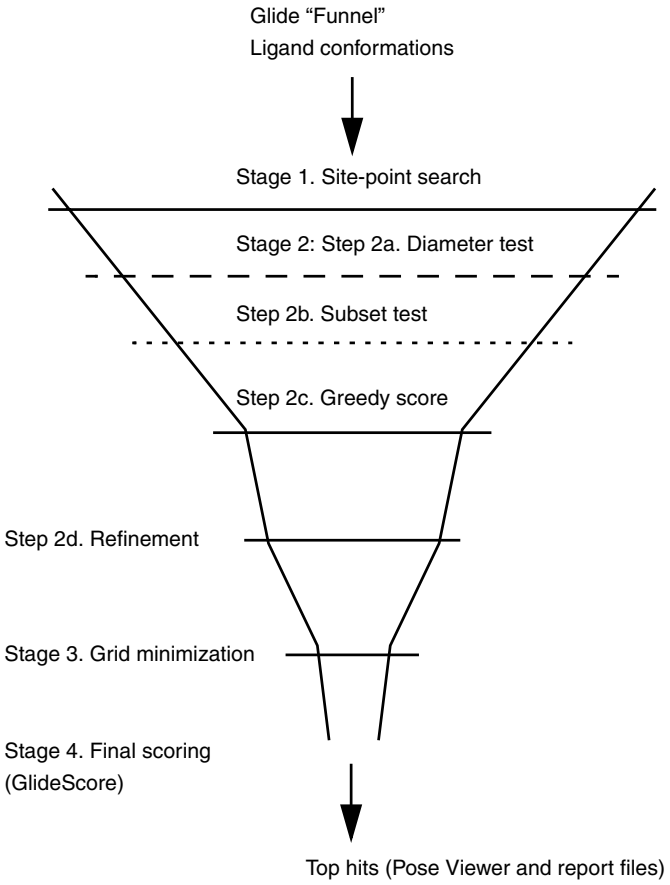


Figure 6.2. The Glide docking hierarchy.

scoring function (Eldridge, et al., *J. Comput. Aided Mol. Des.* **1997,** *11,* 425-445). Much as for ChemScore itself, this algorithm recognizes favorable hydrophobic, hydrogen-bonding, and metal-ligation interactions, and penalizes steric clashes. This stage is called “greedy scoring,” because the actual score for each atom depends not only on its position relative to the receptor but also on the best possible score it could get by moving ±1 Å in X, Y, and/or Z. This is done to mute the sting of the large 2 Å jumps in the site-point/ligand-center positions. The final step in Stage 2 is to re-score the top greedy-scoring poses via a “refinement” procedure (Step 2d), in which the ligand as a whole is allowed to move rigidly by ±1 Å in the Cartesian directions.

Only a small number of the best refined poses (typically 100-400) are passed on to the third stage in the hierarchy—energy minimization on the pre-computed OPLS-AA van der Waals and electrostatic grids for the receptor. The energy minimization typically begins on

a set of van der Waals and electrostatic grids that have been "smoothed" to reduce the large energy and gradient terms that result from too-close interatomic contacts. It finishes on the full-scale OPLS-AA nonbonded energy surface ("annealing"). This energy minimization consists only of rigid-body translations and rotations when external conformations are docked. When conformations are generated internally, however, the optimization also includes torsional motion about the core and end-group rotatable bonds. Unless you specify otherwise, a small number of the top-ranked poses are then subjected to a Monte-Carlo procedure in which alternative local-minima core and rotamer-group torsion angles are examined to try to improve the energy score.

Finally, the minimized poses are re-scored using Schrödinger's proprietary *GlideScore* scoring function. GlideScore is based on ChemScore, but includes a steric-clash term and adds buried polar terms devised by Schrödinger to penalize electrostatic mismatches. The choice of best-docked structure is made using a model energy score (*Emodel*) that combines the energy grid score, the binding affinity predicted by GlideScore, and (for flexible docking) the internal strain energy for the model potential used to direct the conformational-search algorithm. Glide also computes a specially constructed Coulomb-van der Waals interaction-energy score that is formulated to avoid overly rewarding charge-charge interactions at the expense of charge-dipole and dipole-dipole interactions. This score is intended to be more suitable for comparing the binding affinities of different ligands than is the "raw" Coulomb-van der Waals interaction energy. In the final data work-up, you can combine the computed GlideScore and "modified" Coulomb-van der Waals score values to give a composite score that usually helps improve enrichment factors in database screening applications. See the *FirstDiscovery Technical Notes* for more details.

This hierarchical search gives Glide exceptionally high accuracy in predicting the binding mode of the ligand. At the same time, the computational cost is dramatically reduced compared to what would be required for a complete systematic search. The key to this reduction is that the algorithm allows the rotamer groups to be optimized one at a time for a given core conformation and location of the ligand. For example, if there are five rotamer groups and each has three rotamer states, the total number of conformers in the ensemble based on this core conformation/location is $3^5 = 243$. However, if the rotamer groups are optimized one at a time, the number of conformational combinations is only $3 \times 5 = 15$, for a savings of about a factor of 15 in computational effort. While many other time-saving algorithms in Glide contribute to its performance advantages, this fundamental qualitative feature allows large libraries to be screened at an affordable computational cost.

6.1.1 Glide Constraints

A *Glide constraint* is a ligand-receptor interaction requirement. To use Glide constraints, specify up to ten receptor sites for possible ligand interactions when you set up a grid

generation job. When you run a docking job, you can select up to four Glide constraints to apply from the list of receptor constraint sites you defined for that receptor.

In Glide constraint docking jobs, Glide incorporates satisfaction of these constraints into several of its hierarchical filters, allowing prompt rejection of docked poses that fail to meet the requirements. This can significantly speed up docking, improve database enrichment, and in some cases increase docking accuracy.

For information on using Glide constraints, see Section 6.8 on page 94. See the *FirstDiscovery Technical Notes* for details about the method and for a discussion of results.

6.1.2 Glide Extra-Precision Mode

The extra-precision (XP) mode of Glide combines a powerful sampling protocol with the use of a custom scoring function designed to identify ligand poses that would be expected to have unfavorable energies, based on well-known principles of physical chemistry. The presumption is that only active compounds will have available poses that avoid these penalties while at the same time securing a favorable scoring from the terms in the scoring function that reward hydrophobic contact between the protein and the ligand, appropriate hydrogen-bonding interactions, etc. The chief advantages of the XP method are to weed out false positives and to provide a better correlation between good poses and good scores.

Extra-precision mode is a refinement tool designed for use only on good ligand poses. The more extensive XP docking method and specialized XP scoring method are strongly coupled: the more precise poses produced by XP docking are necessary for the more demanding XP scoring method. Because XP docking mode requires more CPU time, you should screen large sets of ligands first in standard-precision mode. Only the top scorers should be docked using XP mode.

Note: If the active site of the complex contains a metal, XP mode should not be used.

For information on using XP mode, see Section 6.3 on page 70. See the *FirstDiscovery Technical Notes* for details about extra-precision mode and for a discussion of results.

6.1.3 Glide/Prime Induced Fit

Glide docking uses the assumption of a rigid receptor. Scaling of van der Waals radii of nonpolar atoms, which decreases penalties for close contacts, can be used to model a slight “give” in the receptor and/or ligand. This may not be sufficient to treat systems where ligand binding induces substantial conformation changes in the receptor (“induced fit.”) Schrödinger has developed a procedure for such cases which uses Prime and Glide to perform induced fit docking. For more about induced fit protocols, see the *Prime User Manual* and *Prime Quick Start Guide*.

6.2 The Glide Panel

Glide searches for favorable interactions between one or more typically small ligand molecules and a typically larger receptor molecule, usually a protein. Each ligand must be a single molecule, while the receptor may include more than one molecule, e.g., a protein and a cofactor. Glide can be run in rigid or flexible docking modes. In flexible docking mode, Glide automatically generates conformations for each input ligand.

To open the Glide panel, choose Glide from the Applications menu in the Maestro main menu bar. The Glide panel has seven tabbed folders:

- Settings
- Site
- Ligand
- Scoring
- Output
- Constraints
- Similarity

6.3 The Settings Folder

The Settings folder defines the basic functionality of the Glide task. The options in this folder determine whether the task is grid generation or ligand docking. For grid generation, this folder contains options for defining the structure in the Workspace, the base name for grid files, and where they will be written. For docking, options in this folder determine the docking mode (standard precision or extra precision), the number-of-atoms and rotatable-bond thresholds for skipping ligands, and which grid files to read.

6.3.1 Glide Function: Set Up Grids or Dock Ligands

Choose Glide function

Using Glide for docking ligands to a receptor is a two-step process. Receptor grids must be generated before ligands can be docked. Glide performs both these functions. The first option in the Settings folder, Choose Glide function, specifies whether Glide is to Set up grids or Dock one or more ligands.

A protein or protein-ligand complex must be included in the Workspace before you set up grid generation. As the name implies, this job only calculates the scoring grids for receptor and saves them to disk. Once this job is completed, subsequent Dock one or more ligands jobs can perform ligand docking using this set of receptor grids.

6.3.2 Docking Mode Options and Using Extra-Precision Mode

Choose docking mode

This menu sets the mode for the Glide docking job. Either Standard speed and precision (SP) or Extra precision (XP) is available. The latter provides better sampling but requires greater CPU time. Generally, you should run your database through the Standard mode first, then take the top 10% to 30% of your final poses and run them through Extra precision, so that you perform the more CPU time-intensive docking simulation on worthwhile poses. The Extra precision option is active only when the Choose Glide function selection is Dock one or more ligands.

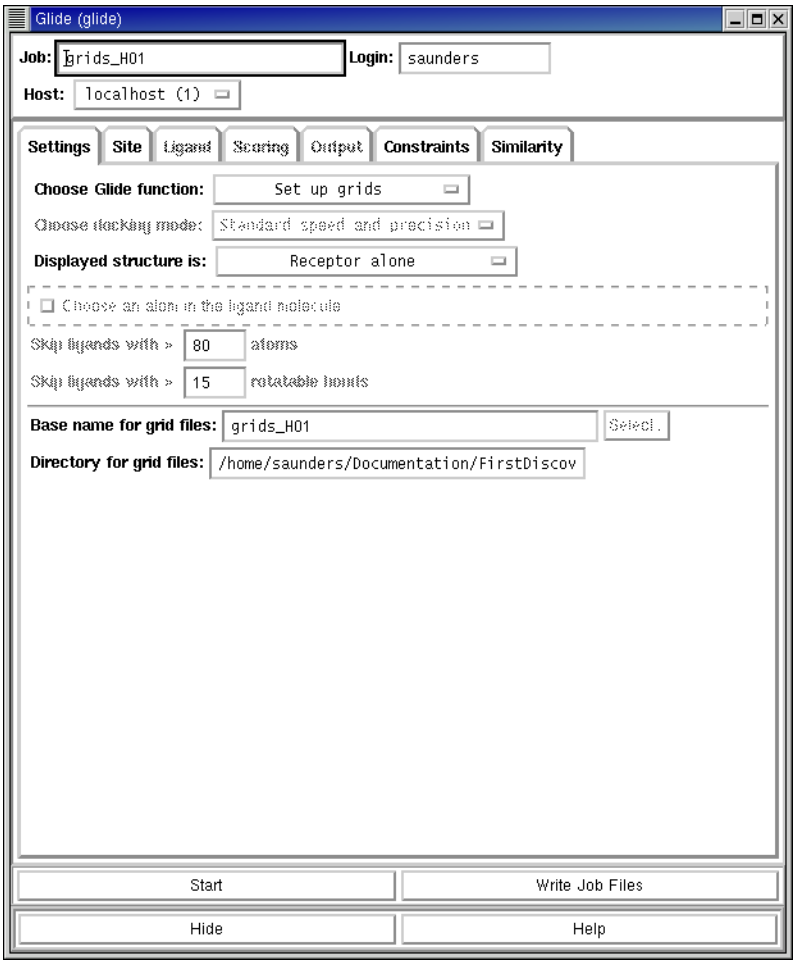


Figure 6.3. The Settings ***folder of the*** Glide ***panel.***

Using Extra-Precision (XP) Mode

Extra-precision mode should be run only on ligand poses that have scored well in a Glide standard-precision run.

Do not use XP mode in systems where the active site contains a metal atom.

To use XP mode:

1. Set up a Glide docking job, choosing docking mode Standard speed and precision, to run on the entire ligand set.
2. In the Output folder of the Glide panel, under Structure output:
   - Select Write ligand pose file *filename*_lib.mae (exclude receptor).
   - Make sure that the default value, 1, is specified in the number field of Keep at most [*n*] poses per ligand.
3. When the standard-precision job is complete, determine which poses are high-scoring enough to be run again using XP mode.

   A rule of thumb is to include the highest-ranking 20% of the poses as docked by SP Glide (i.e., 200 poses for each 1000 database ligands). In some cases, the top 10% may suffice, while in others 30% may be required. Such judgements can be made, however, only if you have known active ligands and can determine their scores and ranks.
4. Extract the selected range of poses from the Maestro output file written by Glide.

   It might be convenient to use the `glide_sort` utility with the `-n` option (number of poses to keep):

   `glide_sort -n` *#poses-to-keep* `-o` *selected-poses-file*
5. Using the ligand poses prepared according to the steps in the previous section, set up a Glide docking job, choosing Extra precision as the docking mode.
6. Complete Glide setup and run the job.

6.3.3 Other Settings Folder Options

Displayed structure is

This option menu has three choices—it is important that the option be selected correctly. If the structure currently displayed in the Workspace is only the receptor protein, without any ligand, set the option menu to Receptor alone.

If the structure currently displayed is the protein and a ligand, choose Receptor plus a single ligand from the option menu. The ligand can be but need not be docked or otherwise used in the Glide setup procedure. If the currently displayed structure is just a ligand to be docked or scored, and this is a Dock one or more ligands job, choose Ligand alone from the option menu. This option is not available for Set up grids jobs. If this is a Dock one or more ligands job, and the contents of the Workspace will not be used in the docking, you can use any of the options.

Choose an atom in the ligand molecule

If you chose Receptor plus a single ligand from the Displayed structure is option menu, then this picking button becomes active. When the button is active, the ligand molecule can be identified by picking any atom in the ligand displayed in the Workspace. The ligand molecule will then be marked with dark green markers. If you chose Ligand alone from the Displayed structure is option menu, Maestro automatically identifies the ligand. For Glide to operate correctly, you must identify the ligand if it is present in the Workspace.

Skip ligands with > n atoms

Use this text box to set the maximum number of atoms for a ligand structure to be docked. Ligand structures in the input file with more than the specified number of atoms will be skipped. The default is 100 atoms. This maximum number of ligand atoms can also be used in the Site folder to calculate the size of the enclosing box for the grids. This number should be reduced if the active-site region is small and enclosed to speed up a docking calculation on a large ligand database.

Skip ligands with > n rotatable bonds

Use this text box to set the maximum number of rotatable bonds a ligand structure may have if it is to be docked. Ligand structures in the input file with more than this number of rotatable bonds will be skipped. The default is 15 rotatable bonds. This number should be reduced if you want only relatively small or rigid ligand hits.

Base name for grid files

This text box specifies the base name *basename* for the grid files. This name is used to find the grid files when existing files are being used for the calculation, or to name the grid files when the grids are being calculated and saved. By default, the base name of the grid files is set to the job name in the Job text box. The default job name is `glidetmp`.

Directory for grid files

This text box specifies a directory that the grid files are to be read from or written to. By default this is set to the current working directory.

6.4 The Site Folder

The Glide Site folder determines where the scoring grids are positioned throughout the receptor and how they are prepared from the structure in the Workspace. The Site folder is enabled only when Choose Glide function in the Settings folder is set to Set up grids. This folder is not used in preparing Dock one or more ligands jobs.

Glide uses two "boxes" to organize the calculation:

- The grids themselves are calculated within the space defined by the purple enclosing box. This is also the box within which all the ligand atoms must be contained.
- Acceptable positions for the ligand center must lie within the green bounding box. This box gives a truer measure of the effective size of the search space. The only requirement on the enclosing box is that it be big enough to contain all ligand atoms when the ligand center is placed at an edge or vertex of the bounding box. Enclosing boxes that are larger than this are not useful: they take up more space on disk and in memory for the scoring grids, which take longer to compute.

The ligand center is defined in a rigid-docking run as the midpoint of the line drawn between the two most widely separated atoms. The definition changes slightly for flexible docking, where the ligand center becomes the midpoint between the two most widely separated atoms of the *core region*—the part of the ligand remaining after each of the end-groups has been stripped off at the "outward" end of the connecting rotatable bond.

The two boxes share a common center. Thus, the operations in the folder that center one box also center the other.

Each rigidly docked ligand or flexibly docked conformation has an associated length, L, which can be defined as twice the distance from the ligand center to the farthest atom. The required relationship between L and the lengths E and B of the enclosing and bounding boxes for successful placement of the ligand center anywhere within the bounding box is:

$$E \geq B + L$$

The enclosing box must be large enough in each dimension to hold the length of the bounding box plus the maximum length of any ligand. If a larger ligand is encountered, not all positions for the center of the ligand in the bounding box are accessible. The effective bounding box for that ligand will be smaller than the dimension nominally specified.

Glide restricts the size of the enclosing box to 50 Å.

6.4.1 Site Folder Features

The Site folder defines parameters for calculating grids. The options in this folder are described below.

Specify center of enclosing box by

These buttons determine how the scoring grids are centered. There are three options:

- The Ligand centroid option centers the grid box at the centroid of the ligand displayed in the Workspace. This is the default option when a ligand has been identified and marked in the Workspace. When you have selected this option, the Specify

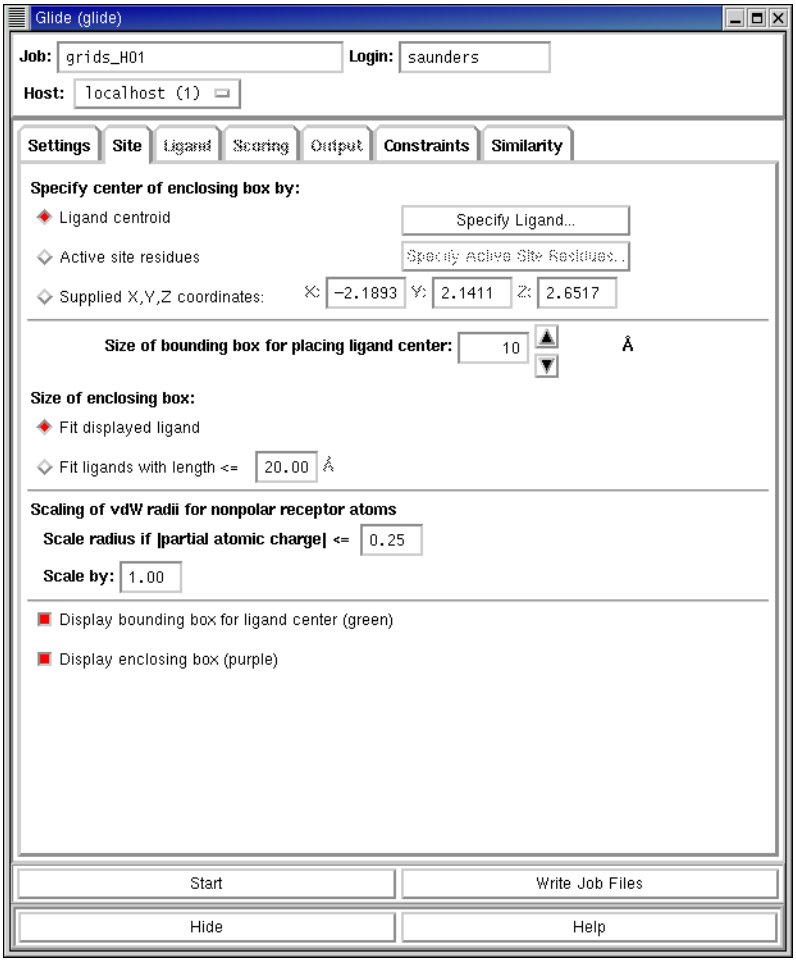


Figure 6.4. The* Site *folder of the* Glide *panel.

Ligand button becomes available. Click it to open the Ligand to Define Grid panel, described in Section 6.4.2. The ligand centroid is computed as half the sum of the smallest and largest x, y, and z coordinates of any atom.

- The Active site residues option centers the box at the centroid of a set of active-site residues. When this option is set, the Specify Active Site Residues button becomes available. Click it to open the Specify Active Site Residues panel, described in Section 6.4.3.
- The Supplied X, Y, Z coordinates option allows you to set the center of the enclosing box directly by typing the coordinates into the X, Y, and Z text boxes. Even if this option is not selected, the values in the X, Y, and Z text fields are updated to show the current box center defined by the selected option. The enclosing box can be repositioned by entering new centering values into the X, Y, and Z text boxes. To provide visual guidance, the X, Y, and Z coordinate axes are displayed at the center of the ligand bounding box.

Size of bounding box for placing ligand center

The up and down arrows increase and decrease the size of the bounding box. The bounding box is the region within the enclosing box within which Glide may place the ligand center. The size of the bounding box can be set to between 6 and 30 Å on each side in increments of 2 Å. If the bounding box is currently displayed in the Workspace, it is redrawn as the box size is changed. The default ligand bounding box is 10 Å on a side.

The "ligand center" used to position the ligand at a site point is different from the Ligand centroid that serves as one choice for positioning the enclosing box and bounding box. For example, in rigid docking runs, the ligand center is taken as the midpoint of a line drawn between the two most widely separated ligand atoms. Example cases suggest that the ligand center and the ligand centroid (computed as the average of the smallest and largest x, y, and z coordinates of any atom) typically differ by 1 – 2 Å. For flexible docking, the difference can easily be greater, for here the ligand center is defined, for each core conformation generated by Glide, as the midpoint of the line between the two mostly widely separated atoms *in the core*. To ensure that the bounding box is not unrealistically small, Maestro sets a minimum value of 10 Å. (Note that the position of the centroid of a set of selected active-site residues can vary even more widely from the ligand center. This is something to keep in mind when selected active-site residues are used to set the box centers.)

Size of enclosing box

As previously noted, the enclosing box is the box that must contain all ligand atoms during the docking procedure. There are two ways to determine the size of this box, but the maximum size is 50 Å on a side.

- The Fit displayed ligand option is available if a receptor and a ligand are displayed in the Workspace and the ligand has been identified. Selecting this option sizes the enclosing box to fit the longest dimension of the displayed ligand plus the bounding box. This option is only appropriate if the ligands to be docked are of the same or smaller size than the displayed ligand.
- The Fit ligands with length <= *n* Å option fits the enclosing box to ligands with the maximum length specified in the text box. The default maximum ligand length is 20 Å. This option is appropriate when you know the maximum ligand size in the input file (or the maximum size to be considered), or when you want to apply a specific restriction on the region within the active site in which any docked ligand must lie.

Scaling of vdW radii for nonpolar receptor atoms

Glide does not allow for flexible receptor docking (for Glide/Prime induced fit docking, see Section 6.1.3 and the Prime documentation), but scaling of van der Waals radii of nonpolar atoms, which decreases penalties for close contacts, can be used to model a slight "give" in the receptor and/or ligand. The options for scaling nonpolar receptor atoms are:

- Scale radius if |partial atomic charge| <= *n*

  Scaling of van der Waals radii is performed only on nonpolar atoms, defined as those for which the absolute value of the partial atomic charge is less than or equal to *n*. The value entered for *n* must be a positive number. The default for atoms in the receptor is 0.25.

- Scale by

  Van der Waals radii of nonpolar receptor atoms are multiplied by this value. The default is not to scale receptor atom radii, and therefore this value is set to 1.00. For ordinary Glide docking, it is recommended that receptor radii be left unchanged, and any scaling be carried out on ligand atoms, as described in Section 6.5 on page 80.

Display bounding box for ligand center (green)

Select this option to display the ligand-center bounding box; deselect it to undisplay the box. The bounding box is drawn as a green wireframe cube in the Workspace. The center of the cube is marked in green by the x-, y-, and z-axes.

Display enclosing box (purple)

Select this option to display the enclosing box; deselect it to undisplay the box. The enclosing box is drawn as a purple wireframe cube in the Workspace.

6.4.2 The Ligand to Define Grid Panel

The Ligand to Define Grid panel is used to select the ligand Maestro uses to center the Glide enclosing box and bounding box. To open the panel, select Ligand centroid in the Site folder and click Specify Ligand.

The selection options of the Ligand to Define Grid panel are described below.

Use displayed ligand

This button requires that the Workspace structure contain a ligand molecule and that the ligand already be identified (e.g., via the Choose an atom in the ligand molecule button on the Settings folder). The displayed ligand is then used to determine the box center.

Use entry name

To select an entry in the Maestro Project Table as the ligand that defines the Glide grid boxes, select this button, and then either type the entry name directly into the adjacent text box or use the Choose selector, which displays a list of all the entries in the current opened project, from which you can select an entry.

Use ligand from external file

To define the Glide grid boxes based on a ligand in an external file, select this option, then use the remaining items in this panel to select a file name and structure number in that file.

Format

Choose the format of the external file from the option menu. The three file formats listed below are supported for reading in ligand files from local or network-mounted disks. Glide does not support Mol2 formatted files.

- Maestro: Maestro-written files (extensions `.mae`, `.out`, or `.dat`)
- MDL SD: SD-formatted files (extensions `.mol` for single structure files and `.sdf` for multiple structure files)

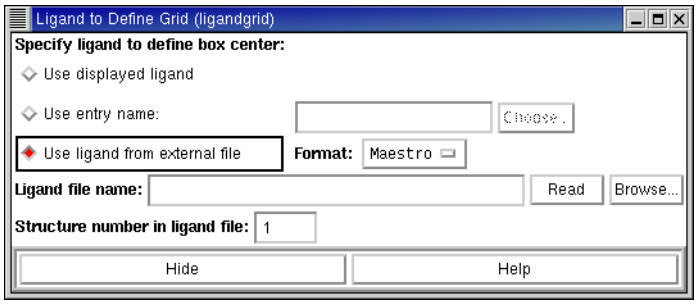


Figure 6.5. The Ligand to Define Grid ***panel.***

- PDB: Rutgers Center for Structural Biology Protein Data Bank files (extensions `.pdb` or `.ent`)

Ligand file name

This text box allows you to specify the location of the ligand if it is not already loaded into the Workspace. There are two buttons to the right of the text box: Read and Browse. If there is no displayed ligand, the selected ligand is added to the Workspace. Click Browse to browse for the ligand file, or enter the file name in the text box and click Read.

Structure number in ligand file

If the ligand file contains multiple structures, you can specify the structure to be used as the ligand by entering the number that corresponds to its position in the file. The default is to use structure 1.

6.4.3 The Active Site Residues Panel

Use the Active Site Residues panel to define selected active-site residues of a receptor for a Glide calculation. The centroid of the selected residues is then used to center the Glide grid boxes. To open the panel, select Active site residues in the Site folder and click Specify Active Site Residues.

The selection options of the Active Site Residues panel are described below.

Active-site residues list

The text area in the upper portion of the panel lists the active-site residues that are currently defined. The list keeps track of the selections made and displays them. Items displayed in this list can be selected for deletion (see below).

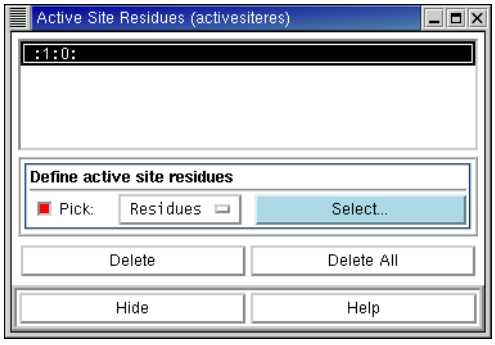


Figure 6.6. The Active Site Residues ***panel.***

Define active site residues

The picking controls in this section of the panel allow you to select the residues that constitute the active site. For more information about picking controls and the Atom Selection dialog box (ASD), see Section 2.6 on page 21, search for "picking controls" in the Maestro online help, or see the *Maestro User Manual*.

6.5 The Ligand Folder

The Glide Ligand folder allows you to identify the ligands to be docked or scored by Glide. You can also specify a reference ligand for use in rms comparisons of docking accuracy.

6.5.1 Ligand Folder Features

Dock displayed ligand

When this button is on, the ligand currently displayed in the Workspace will be docked by Glide. This option is enabled only when the ligand molecule has been identified in the Workspace (see the Displayed structure is option menu in the Settings folder in Section 6.3.3 on page 72).

Dock displayed ligand only

If you have selected Dock displayed ligand, this option becomes available. Choose it if you do not want to dock any ligands from the Project Table or external files.

Dock selected entries from project table

This button indicates that Glide should dock all the entries selected in the Project Table. This can be done instead of docking ligands from external files, but with or without docking the displayed ligand.

Dock ligand(s) from files

When this button is on, one or more ligands will be read from one or more structure files and docked by Glide instead of, or in addition to, the currently displayed ligand. Using multiple files is supported only for PDB format, and each PDB file must describe one ligand. Maestro and SD file formats allow multiple structures to be included in one file, but only one Maestro or SD file can be specified here.

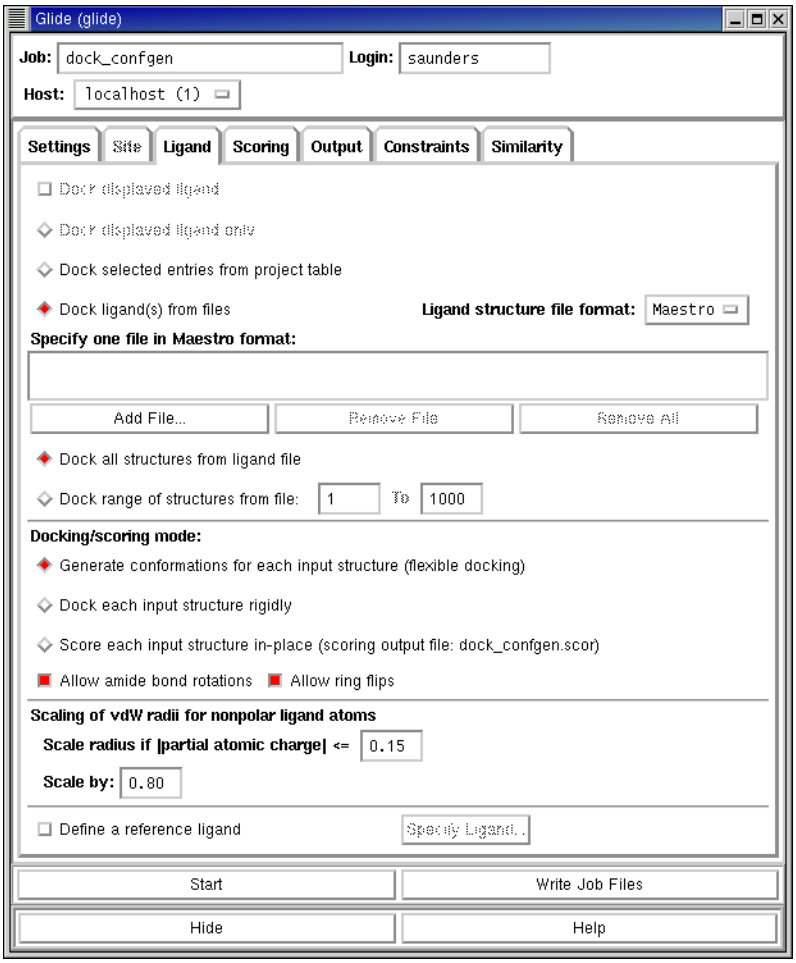


Figure 6.7. The* Ligand *folder of the* Glide *panel.

Ligand structure file format

Glide supports the three file formats listed below for reading ligand files. Glide does not support Mol2 formatted files. Note that Glide automatically skips fragmented ligands (e.g., salts with counterions present), ligands containing lone pairs, and ligands containing unparameterized elements (e.g., arsenic).

- Maestro: Maestro-written files (extensions `.mae`, `.out`, or `.dat`). One file may be specified.
- MDL SD: SD-formatted files (extensions `.mol` for single structure files and `.sdf` for multiple structure files). One file may be specified.

- PDB: Rutgers Center for Structural Biology Protein Data Bank files (extensions `.pdb` or `.ent`). When this option is chosen, more than one file can be specified.

Specify one [or more] file[s] in *format* format

This list specifies the ligand files Glide will dock (listed in order of user entry when PDB format is used). The file formats given in the Ligand structure file format option menu are supported.

To add a file, click Add File. The Add ligand structure files panel is displayed. In this panel ligand files can be selected and opened for placement in the list.

The two buttons to the right of the Add File button allow you to remove files already placed in the list. The Remove File button removes a highlighted ligand file and the Remove All button removes all listed ligand files.

For PDB format, any number of ligand files may be specified. For other formats, only a single file may be specified (the Add File button is disabled after the first file is added), but the file may contain any number of ligand structures.

Dock all structures from ligand file/Dock range of structures from file

When multiple ligands are read from a single Maestro or SD file, Glide will Dock all structures from ligand file by default. Selecting the Dock range of structures from file option instead allows a subset of consecutive structures to be docked, numbered as they are listed in the file.

This facility provides a convenient mechanism for dividing a large input ligand file into subsets for concurrent processing on a multi-processor machine—provided you have sufficient licenses to run multiple copies of Glide simultaneously. For example, you could start a first calculation by clicking the Start button with `1` and `100` entered in the text boxes. Then, you could enter `101` and `200`, change the job name in the Job text box (to avoid overwriting output files), and click the Start button again to dispatch a second group of ligands. Continuing in this way, you could quickly submit as many sub-jobs as are needed to cover the range of ligands to be docked (subject to licensing restrictions).

Such multiple submissions, however, are more commonly—and more conveniently—handled by selecting Write Job Files to write out a single “template” input file. You can then use this template file as input to the `para_glide` script, which creates (and if specified, submits for distributed processing) several smaller subjobs covering the range of ligands in the starting template file you specify. The `para_glide` script is installed in the `$SCHRODINGER/utilities` directory (see the usage summary on page 110). Also see the *First Discovery Technical Notes* for more details.

Other uses for range bounds settings:

- To dock all structures from the first range value to the end of the file, set the To value to `0`. If the first value is `1` and the second value is `0`, all structures in the file will be docked.

  Note: The To (second) value should never be less than the from (first) value, except when To is set to `0`.

  Note: Neither value can be less than zero.

- In the event that a problem with Glide or with the input ligand causes a submitted Glide job to terminate abnormally, you can set these range bounds to pick up after the point in the input ligand file at which the problem occurred.

Docking/scoring mode

The docking/scoring modes are:

- Generate conformations for each input structure (flexible docking)
- Dock each input structure rigidly
- Score each input structure in-place (scoring output file: *jobname*.scor)

These options determine whether input ligands are to be docked flexibly, rigidly, or not at all (evaluated and scored "in-place"). For flexible docking, the options here further determine whether amide bond rotations are to be penalized and whether ring conformation flips will be allowed. Score in-place options include whether to count amide bonds as rotatable.

Generate conformations for each input structure (flexible docking)

This option directs Glide to generate conformations internally during the docking process, a procedure known as *flexible docking*. At present, conformation generation is limited to variation around acyclic torsion bonds, generation of conformers for five– and six–membered rings, and generation of pyramidalizations at certain trigonal nitrogen centers (e.g., in sulfonamides). To carry out more complete conformational exploration, you would have to explicitly include in the input ligand file a representative structure for each larger-ring conformation believed to be relevant.

When flexible docking is selected, the amide and ring conformation options are:

- Penalize amide bond rotations (SP: freeze; XP: restrain to cis/trans)

  The default is *not* to penalize amide bond rotations.

  In SP (default) docking, selecting this option freezes amide bonds in their input conformation throughout docking; no rotations will occur.

In extra-precision (XP) docking, selecting this option does not prevent *cis* to *trans* or *trans* to *cis* conformation changes, but non-planar amide bonds are heavily penalized.

- Allow ring flips

  The default is to allow ring flips. Deselect this option if you want nonaromatic five– and six–membered rings to remain in their input conformation throughout docking.

Dock each input structure rigidly

"Rigid" docking allows the existing ligand structure to be adjusted, but the conformation generation step is skipped. Since conformations are not generated, the amide and ring conformation options are not available.

Score each input structure in-place (scoring output file: *jobname*.scor)

Choose this option if you want to use the input ligand coordinates to position each ligand for scoring. (This option performs scoring but not docking, since other ligand positions are not sampled.) One use of this capability is to score the cocrystallized ligand in its original or modeled position. Another use is to post-process ligand poses generated by Glide to obtain additional components of the GlideScore (or energy score) prediction of the binding affinity. The output is written to a file named *jobname*`.scor`, where *jobname* is taken from the Job text box found at the upper left corner of the panel. As in other docking options, you can score the displayed ligand and/or one or more ligands read from an external file.

- Do not count amide bonds as rotatable

  This option is available when the score in-place option is selected. The default is to include amide bonds in the total rotatable bonds count. Select this option to exclude amide bonds from the total.

Scaling of vdW radii for nonpolar ligand atoms

As previously noted, Glide does not allow for flexible receptor docking (for induced fit docking, see Section 6.1.3), but to model a bit of give in the receptor, you can scale the van der Waals radii of nonpolar atoms in the ligand and/or receptor, causing lower penalties for close contacts.

Scaling of nonpolar receptor atoms must be set up in the Glide Site folder before grid generation is performed. Use the Ligand folder options described here to use vdW scaling of nonpolar ligand atoms in docking.

The Scale radius if |partial atomic charge| option defines which atoms are considered sufficiently nonpolar to be subject to scaling. This is the absolute value of the partial charge, so the number entered must be positive. The default cutoff for ligand atoms in Glide 3.0 is a

partial charge with absolute value smaller than or equal to 0.15 electrons. The Scale by text box sets the scaling factor. The default is 0.80. To turn van der Waals scaling of ligand atom radii off, set the scale factor to 1.0.

Define a reference ligand

When selected, this option allows you to specify the conformation, position, and orientation of a reference ligand for Glide to use to assess the accuracy of docking topologically identical ligands. Glide reports the rms distance between nonhydrogen atoms of the docked ligand and the reference ligand when the docked ligand and the reference ligand are recognized as being conformers (this requires that they be topologically equivalent and that the ordering of corresponding atoms be the same). When this option is on, the Specify Ligand button will be active, enabling you to open the Reference Ligand panel.

6.5.2 The Reference Ligand Panel

The Reference Ligand panel defines the reference ligand Glide uses to compute rms comparisons to conformationally-related docked ligands. Note that if you want a calculated rms for the reference ligand itself, make sure it is first in the ligand list.

To open the Reference Ligand panel, select Define a reference ligand in the Ligand folder and click Specify Ligand. The selection options of the Reference Ligand panel are described below.

Use displayed ligand

This button requires that the Workspace structure contain a ligand *and* that the ligand already be identified (e.g., via the Choose an atom in the ligand molecule button in the Glide Settings folder). The displayed ligand is then used as the reference ligand.

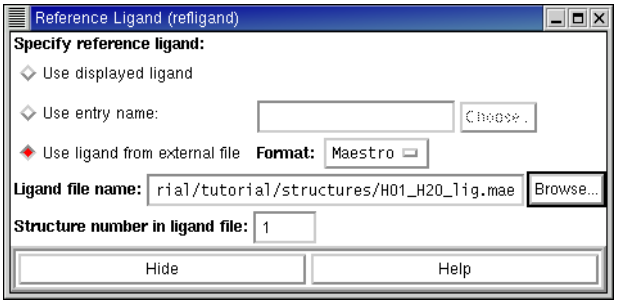


Figure 6.8. The Reference Ligand ***panel.***

Use entry name

To select an entry in the Project Table as the reference ligand, select this button, and then either type the entry name directly into the adjacent text box, or use the Choose selector, which displays a list of all the entries in the current opened project, from which you can select an entry.

Use ligand from external file

Format

The three file formats listed below are supported for reading in ligand files from local or network-mounted disks. Glide does not support Mol2 formatted files.

- Maestro: Maestro-written files (extensions `.mae`, `.out`, or `.dat`)
- MDL SD: SD-formatted files (extensions `.mol` for single structure files and `.sdf` for multiple structure files)
- PDB: Rutgers Center for Structural Biology Protein Data Bank files (extensions `.pdb` or `.ent`)

Ligand file name

Specify the name of the file in which to find the reference ligand. The Browse button to the right of the text box can be used to locate the file.

Structure number in ligand file

If the ligand file contains multiple structures, you can specify the structure to be read as the reference ligand by entering the structure number of the ligand in the file.

6.6 The Scoring Folder

The settings in the Scoring folder define the manner in which Glide processes the ligand poses (i.e., combinations of position, orientation, and conformation) it generates. These poses pass through a series of hierarchical filters that evaluate the interaction of the ligand with the receptor. Stages 1 and 2 (see Figure 6.2 on page 67) test the spatial fit of the ligand to the defined active site, and examine the complementarity of ligand-receptor interactions using a grid-based method patterned after the empirical ChemScore function (Eldridge, et al., *J. Comput.-Aided Mol. Des.* **1997,** *11,* 425-445). Poses that pass these initial screens enter the final stage of the algorithm, which involves evaluation and minimization of a grid approximation to the OPLS-AA nonbonded ligand-receptor interaction energy.

Final scoring, which by default uses Schrödinger's proprietary GlideScore multi-ligand scoring function, is then carried out on the energy-minimized poses. Finally, if GlideScore

was selected as the scoring function, a composite “Emodel” score is used to rank the poses and to select the pose or poses to be reported to you. Emodel combines GlideScore, the nonbonded interaction energy, and, for flexible docking, the excess internal energy of the generated ligand conformation.

There are three main sections in the Scoring folder:

- Refinement of initial poses on coarse grid
- Energy Minimization
- Final Scoring

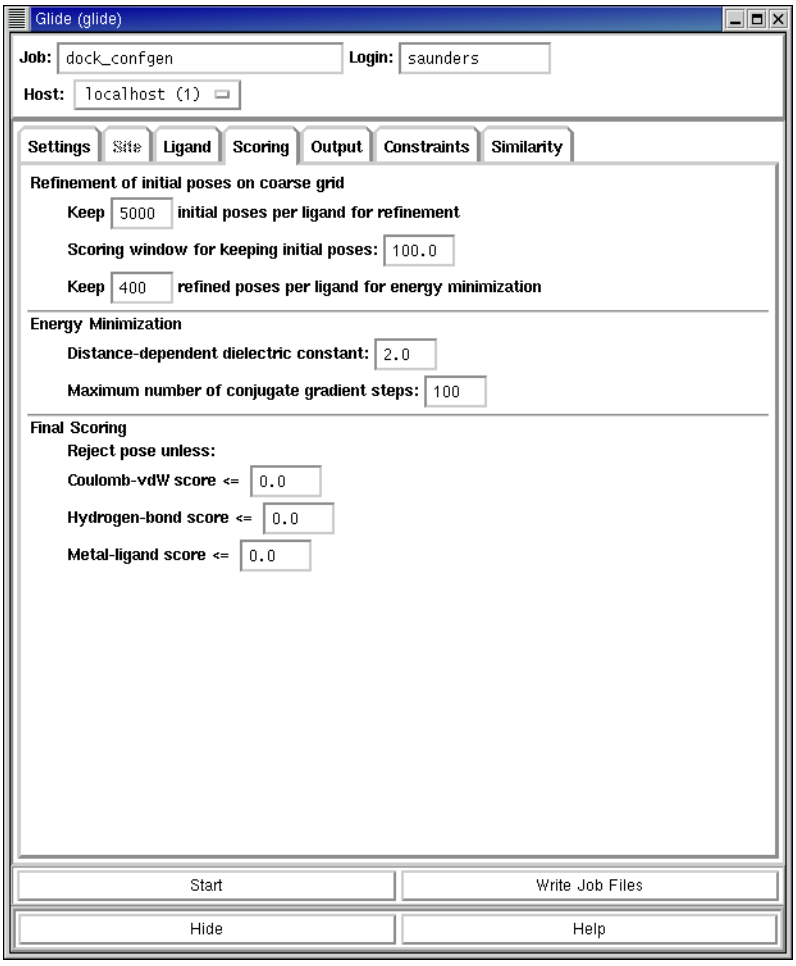


Figure 6.9. The* Scoring *folder of the* Glide *panel.

6.6.1 Refinement of Initial Poses Section

The Refinement of initial poses on coarse grid section of the Scoring folder tailors the way poses pass through the filters for the initial geometric and complementarity "fit" between the ligand and receptor molecules. The grids for Stage 2 of the hierarchy contain values of a scoring function representing how favorable or unfavorable it would be to place ligand atoms of specified general types in given elementary cubes of the grid. These grids have a constant spacing of 1 Å. The *rough score* for a given pose of the ligand relative to the receptor is simply the sum of the appropriate grid scores for each of its atoms. Negative scores are favorable, so the lower (more negative) the better.

The initial "rough scoring" is done on the "coarse grid," on which the possible positions for placing the ligand center are separated by 2 Å (twice the elementary cube spacing) in X, Y, and Z. The "refinement" step rescores the (successful) rough-score poses after the rigid translational repositioning of –1, 0, or +1 Å in X, Y, and Z that gives the repositioned ligand the best possible score. This procedure effectively doubles the resolution of the scoring screen.

The three text boxes in this section are described below.

Keep ___ initial poses per ligand for refinement

This text box sets the maximum number of poses per ligand to pass to the grid-refinement calculation. The default maximum number depends on the type of docking specified:

- For flexible docking jobs in general, the default is 5000 poses.
- If Glide Constraints have been applied to a flexible docking job, 500 poses.
- If extra-precision (XP) docking has been selected, the maximum number of poses is internally adjusted to a minimum of 800.
- For rigid docking, the default is 1000 poses. This number is not changed by the application of Glide Constraints.

You can change the default by entering any integer value greater than zero.

Scoring window for keeping initial poses

This text box sets the rough-score cutoff for keeping poses for refinement. To survive, the score of a given pose must be within the value entered of the best pose accumulated so far. The default setting is 100.0, but you are allowed to choose any real value greater than zero.

Keep ___ refined poses per ligand for energy minimization

This value allows at most the number of poses specified per ligand to be energy minimized on the OPLS-AA nonbonded-interaction grid. The default number depends on the type of docking specified:

- For flexible docking jobs in general, the default is 400 poses.
- If Glide Constraints have been applied to a flexible docking job, 40 poses.
- If extra-precision (XP) docking has been selected, the number of poses is internally adjusted to a minimum of 800.
- For rigid docking, the default is 100 poses. This number is not changed by the application of Glide Constraints.

The range for this setting is 1 to the value in the Keep ___ initial poses per ligand for refinement text box.

6.6.2 Energy Minimization Section

The third stage of the Glide algorithm (see Figure 6.2 on page 67) evaluates and minimizes poses that survive the Refinement Of Initial Poses On Coarse Grid scoring phase.

The selection options in this section are described below.

Distance-dependent dielectric constant

Glide uses a distance-dependent dielectric model, in which the effective dielectric "constant" is the supplied constant multiplied by the distance between the interacting pair of atoms. This text box allows you to set the dielectric constant to any real value greater than or equal to 1.0. The default setting is 2.0. Do not change this setting because Glide's sampling algorithms are optimized for this value.

Maximum number of conjugate gradient steps

This text box allows you to alter the maximum number of minimization steps used by the conjugate-gradient minimization algorithm. The default number of steps is 100, but you can choose any value greater than or equal to 0. A "minimization" of 0 steps does a single-point (current energy) calculation on each pose that survives rough-score screening, or on the input pose if no screening was done. If the Score each input structure in-place option in the Ligand folder is selected, the input pose is used.

6.6.3 Final Scoring Section

The controls in this section are related to the GlideScore method used in the final scoring of the poses. The default is to reject a pose unless:

Coulomb-vdW score <=

If the pose has a Coulomb-van der Waals score greater (more positive) than this value, the pose is rejected. The default value is 0.0 kcal/mol. This means that poses that interact favorably with the protein site, however weakly, are retained, whereas poses that interact unfavorably are rejected. For Glide 3.0, you should not change this value. An exception might be for exploratory runs with actives that you know bind to the (flexible) receptor, if you want to find out how well or poorly Glide's rigid-receptor docking model treats them. In such cases, you might choose a large positive value, such as 20,000, to make sure that you can "track" all of the known actives.

Hydrogen-bond score <=

In the current version of GlideScore, each hydrogen bond with an appropriate H…X distance and an appropriate Z-H…X angle receives a score of about –0.9 kcal/mol. If the pose has a hydrogen-bond score greater (more positive) than the value specified, the pose is rejected. The default value is 0.0. As hydrogen-bond scores can only be negative, use of the default value accepts all poses. See Eldridge, et al. (*J. Comput.-Aided Mol. Des.* **1997,** *11,* 425-445) for a definition of the ChemScore Hbond term that GlideScore uses.

Metal-ligand score <=

If the pose has a metal-ligand score greater than (more positive) or equal to this value, the pose is rejected. The default value is 0.0. For Glide 3.0, you should not change this value.

6.7 The Output Folder

The Glide Output folder organizes various options that control the final output of poses that pass successfully through Glide's various scoring stages. There are two main sections to the Output folder:

- Elimination of duplicate poses (clustering)
- Structure output

6.7.1 Elimination of Duplicate Poses Section

This option ensures that poses found to be identical (within the specified criteria) are not replicated in the Glide output. The default settings are:

Discard pose as duplicate if:

RMS distance < 0.5 Å and maximum atomic displacement < 1.3 Å

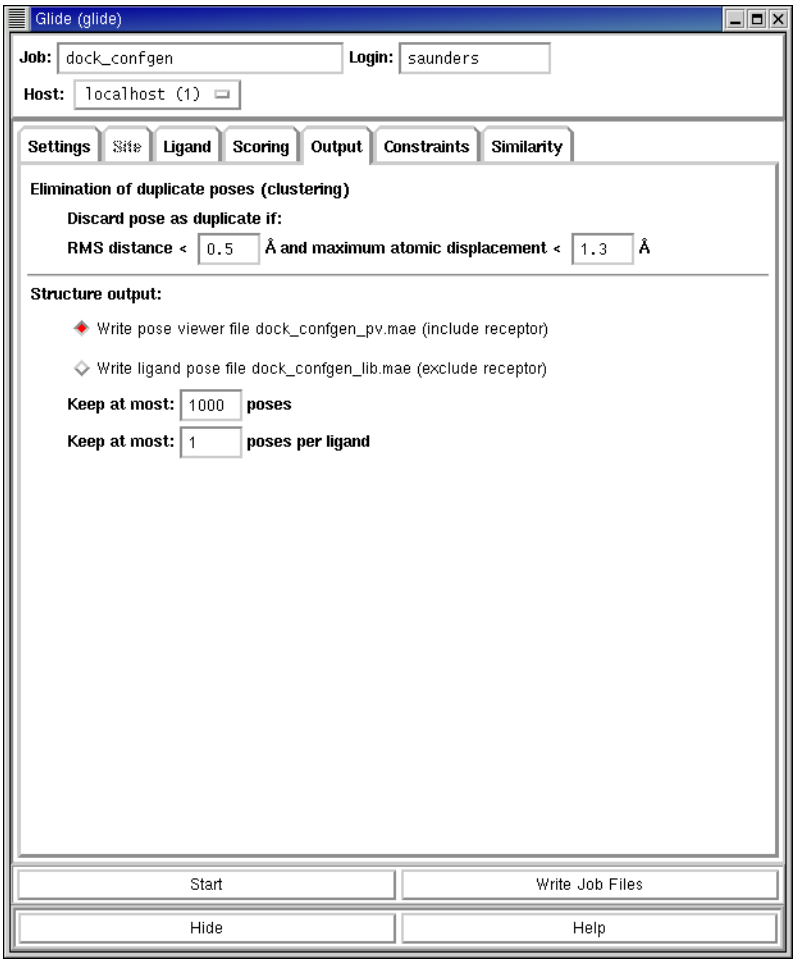


Figure 6.10. The* Output *folder of the* Glide *panel.

from a pose previously selected for inclusion in the reported output. You can choose any values greater than 0.0 Å. The rms distance is based on a comparison of heavy atom (non-hydrogen) coordinates.

6.7.2 Structure Output Section

The options in this section are described below.

Structure output

The final list of poses that pass Glide's criteria is written to a multi-structure Maestro file. The options are:

- Write pose viewer file *jobname*_pv.mae.

  This option includes the receptor structure in the output Maestro file *jobname*_pv.mae, where *jobname* is taken from the Job text box in the upper left of the panel. This version of the Pose File is intended for use with Maestro's Glide Pose Viewer facility.

- Write ligand pose file *jobname*_lib.mae.

  This option causes only ligand structures to be written to the output Maestro file *jobname*_lib.mae, where *jobname* is taken from the Job text box found at the upper left of the panel. This file, also called a "Ligand Library" file, cannot be used by the Pose Viewer, but might be appropriate if the output poses are intended for input to a subsequent Glide job or for some other purpose.

  A report file that lists critical information about the scoring of the retained poses is also written. Like the output pose file, the entries in the report file are ordered by the selected final-scoring function (e.g., by GlideScore or by the modified Coulomb-van der Waals energy score). The file name is *jobname*.rept.

Keep at most ___ poses

This text box limits the total number of the predicted best-binding poses written to the output file. The default value is 1000 poses. This number should be increased in large-scale database screens if substantially more than 1000 ligands are docked in individual jobs.

Keep at most ___ poses per ligand

This text box limits the number of poses per ligand that will be written to the output file. The default choice of 1 pose per ligand is intended for use in database screening applications. A larger value may be appropriate for lead-optimization studies, or whenever several "reasonable" poses for an individual ligand are wanted—for example, to generate a variety of docked poses for study by Liaison or another post-docking program.

6.7.3 Raw Pose Files, Output Pose Files, and `glide_sort`

In flexible docking runs (but not rigid docking runs), Glide continually appends good poses to an intermediate *jobname*_raw.mae pose file during the docking calculation. The "raw" in the file name indicates that these poses have not been sorted by score. The *jobname*_raw.mae file has the same format as the *jobname*_lib.mae or

*jobname*_pv.mae file. This file contains all poses that qualify for inclusion in the final report (i.e., subject to the Keep at most specifications) at the current stage of the calculation.

At the successful conclusion of a flexible docking job, Glide invokes the glide_sort utility. This utility filters the overcomplete structure set in the raw pose file using the Final Output filters from the Scoring folder (see Section 6.6 on page 86) and the duplicate poses filter from the Output folder. It then sorts the filtered poses by score (the default is to sort using the GlideScore) to produce the output pose file, either *jobname*_lib.mae or *jobname*_pv.mae. A report file (*jobname*.rept) is also written.

The default glide_sort settings usually suffice to produce a sorted and filtered output file containing all the useful poses that were in the raw file. Once the output file has been written, the *jobname*_raw.mae file can be deleted to save space.

You can also run glide_sort from the command line, either during or after a Glide docking job, on raw or output pose files. (Like other Glide utilities, glide_sort is installed in your $SCHRODINGER/utilities directory. See Section 6.11 for a usage summary.) Running glide_sort yourself is useful in a number of situations:

- If you decide, after the output file has been produced, to sort poses using different criteria, you can run glide_sort with different settings on your output *jobname*_lib.mae or *jobname* _pv.mae file.
- If you want to use a custom scoring function, you can specify it by setting coefficients for the component energy scores, creating a new composite scoring function.
- If Glide stops unexpectedly, you can run glide_sort on the *jobname*_raw.mae file to recover all relevant data generated to that point.
- If you want an assessment of progress made so far, you can run glide_sort on the *jobname*_raw.mae file at any time during the docking run.

In rigid docking runs, there is no *jobname*_raw.mae file produced. Therefore, glide_sort cannot be used to recover structures that were rejected during the docking calculation. It is still possible to whittle down the *jobname*.rept file further by using glide_sort on the output .mae file instead, if you want to apply stricter criteria for saving poses than in the original run.

For in-place scoring, there is no structure output file, since the input structures are not altered. Instead of the .rept file, a .scor file is produced, containing scoring information for the input structure.

6.8 The Constraints Folder

The Constraints folder is used to set up docking constraints: receptor-ligand hydrogen bond or metal-ligand interactions that you think are important to the binding mode. By setting such prerequisites, often you can significantly enrich the final results and speed up docking, as Glide is able to discard ligands, conformations, or poses that do not meet these criteria early on in their evaluation for docking suitability.

6.8.1 Using Glide Constraints

To use Glide constraints, you must perform both of these steps:

1. Define Glide constraints during grid generation setup.
2. Apply Glide constraints during docking setup.

Any Glide constraints you would like to apply to docking must be defined when the receptor grids are generated.

When you are specifying options for the Set up grids job, open the Glide Constraints folder and select atoms in the receptor which you would like to interact with the ligand during docking. It may be helpful to undisplay most of the receptor, leaving only residues within a short distance of the ligand visible.

As it is picked, each receptor atom is marked with a red cross and padlock in the Workspace. Receptor atoms that are symmetry-equivalent to the one picked are part of the same constraint and are marked along with the picked atom. You can define up to ten different Glide constraints for a single grid generation job. As well as being marked in the Workspace, they are listed in the Glide Constraints folder Receptor constraint sites box. Symmetry-equivalent atoms appear in square brackets.

When you set up a docking job using the grids you have generated, the Glide constraint sites associated with those grids appear in the Receptor constraint sites list in the Glide Constraints folder. For any single docking job, you can select up to four Glide constraints. Each site you select will be required to interact with the ligand during docking. Select from one to four sites for constraints by clicking on them to highlight them. To deselect a constraint, click it again.

If there are no constraints selected when you start the docking job, no constraints will be applied. Multiple docking jobs can be run using the same receptor grids but choosing different subsets of the defined Glide constraints.

Note: Grid files containing Glide constraint definitions that were generated in FirstDiscovery versions earlier than 3.0 will still work in FirstDiscovery 3.0, but you will not be able to use the subset feature. You can use all of the constraints (the default) or use none of

them. The latter option requires you to edit the input file. See the *FirstDiscovery Release Notes* for instructions.

6.8.2 Glide Constraints Folder Features: Grid Generation

When the Glide task chosen in the Settings folder is Set up grids, the following features appear in the Constraints folder:

Receptor constraint sites

This list remains empty until you have selected a site in the receptor for a Glide constraint. Each receptor atom chosen is identified by atom number: atom type: residue type and number: chain. Symmetry-equivalent atoms defined by a single constraint are separated by commas within square brackets.

Pick to choose atoms

This option is selected by default, allowing you to define possible Glide constraints by picking atoms in the receptor.

Show markers

Selected by default, this option marks the receptor constraint sites you select with red crosses and padlocks.

Up to ten symmetry-distinct receptor atoms can be chosen as possible constraint sites, though no more than four constraints may be applied for a given docking job. Typically, these are receptor atoms that are known to be important in binding from structural or biochemical data. But they can be any receptor atom capable of forming a hydrogen bond with the ligand, or any metal ion included in the receptor.

In the case of a hydrogen-bonding interaction, the receptor atom must be a polar hydrogen, nitrogen, or oxygen. If you choose an atom with one or more symmetry-equivalent atoms in its functional group, the symmetry-equivalent atoms will all be selected as well, and collectively count as one constraint. For example, if you create a constraint by picking one oxygen atom of a carboxylate group, Glide includes the other oxygen atom in the same constraint. A ligand interaction with either oxygen atom will satisfy that single constraint.

In the case of a metal-ligand interaction constraint, the receptor atom must be a metal ion. Metal-ligand constraints can also include restrictions on the formal charges of the interacting ligand atoms. These requirements are added during setup of docking jobs.

The receptor atoms selected must also be close enough to the ligand to make satisfying the constraints feasible. You do not need to specify any individual distances or angles between receptor and ligand atoms. The appropriate bounding values for these measurements are

part of Glide's chemical knowledge. Distance requirements are incorporated using the enclosing box for the ligand, the region in which ligand atoms may be found. When the enclosing box has been defined, Maestro displays the box in purple. The receptor atoms selected for constraints must be inside this purple enclosing box or within bonding range of it.

When constraint setup is complete and the grid generation job is run, Glide will write a file containing the information about the specified receptor atoms (their locations and types). Subsequent docking jobs will use this file in order to determine whether a given ligand pose satisfies the constraints. If the base name for writing grid files is *gridbase*, then the file will be called *gridbase*`.cons`. Glide docking will get the constraints information it requires from *gridbase*`.cons`.

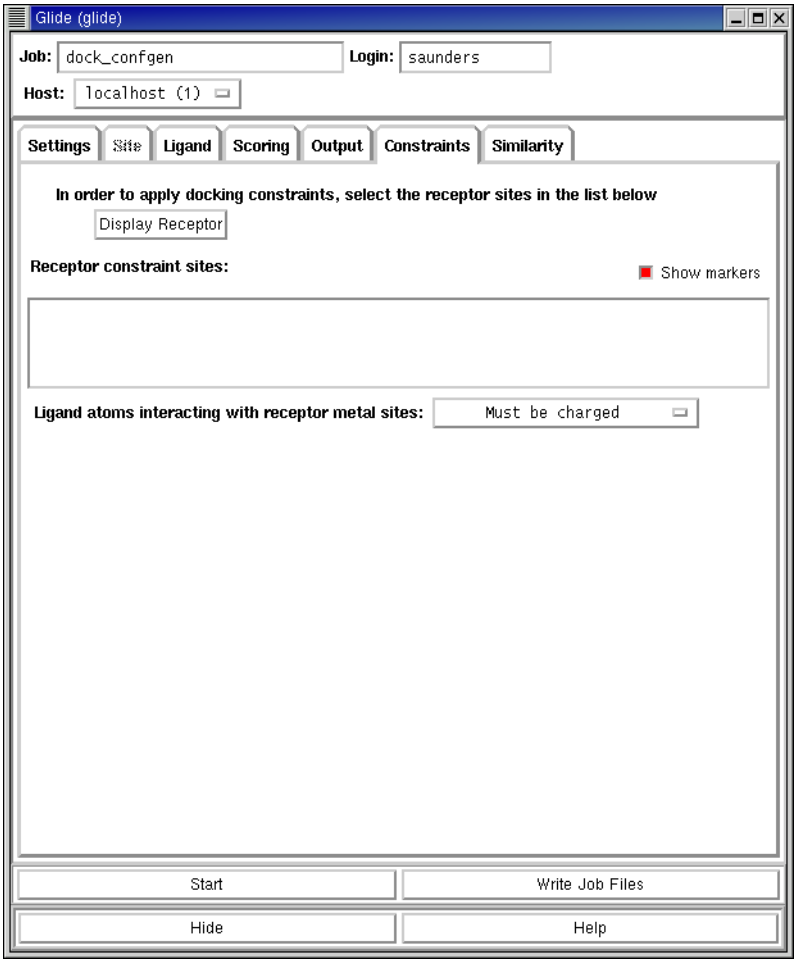


Figure 6.11. The Constraints ***folder of the*** Glide ***panel***

6.8.3 Glide Constraints Folder Features: Docking

Once you define Glide constraints in grid generation, you can apply these constraints to docking jobs. During setup of the docking job, return to the Constraints folder and select up to four sites from the list. The same set of grids can be used for docking jobs with or without constraints. The following features appear in the Constraints folder when Dock one or more ligands has been selected in the Settings folder:

Display Receptor

Click this button to display the receptor in the Workspace. If Show markers has not been deselected, every receptor constraint site defined in grid generation will be marked.

Show markers

Selected by default, this option marks possible receptor constraint sites with red crosses and padlocks. You must still select a constraint from the list to employ it during docking.

Receptor constraint sites

The Glide constraints defined in the receptor grids are listed in this box. Each atom is identified by atom number, atom type, residue type and number, and chain, separated by colons. Symmetry-equivalent atoms defined by a single constraint are separated by commas within square brackets. Select (highlight) the constraints to be used in docking. If none of the constraints in the list are selected, the docking job will proceed without using Glide constraints.

Ligand atoms interacting with receptor metal sites

For a docking job with Glide constraints involving metal-binding interactions, you must provide one more piece of information. You need to choose whether ligand atoms will be required to have a specified formal charge in order to be considered as satisfying constraints. The choices are: Must be charged (the default, and usually the best choice), Must be neutral, or May be either charged or neutral. Glide will apply the requirement you choose for binding to all receptor metal ions you have selected for constraints. Note that for a ligand coordinating group such as a carboxylate, the two oxygens are regarded as having non-zero formal charges of –0.5. Therefore either oxygen can be used to satisfy the constraint, provided that Must be charged or May be charged or neutral is selected.

6.9 The Similarity Folder

6.9.1 Introduction to Similarity Scoring in Glide

Similarity scoring is a means of quantifying how much alike or different two molecules are. In Glide, similarity scoring enables you to reward or penalize ligands based on how they compare to specified sets of active or inactive molecules. Including similarity scoring in Glide docking jobs can yield significantly improved enrichment factors in database screening experiments.

Glide determines the similarity of two molecules by comparing lists of their atom pairs (AP). An atom pair is defined by two atoms and the number of bonds in the shortest path connecting them. For similarity scoring purposes, each atom is characterized by an atom type that depends on its atomic number, neighbors, bond orders, and formal charges. (These atom types are different from those associated with force fields, used elsewhere in FirstDiscovery). Two molecules with no APs in common are assigned a similarity score of 0.0, and two with identical AP lists have a similarity of 1.0.

Using similarity scoring in a Glide docking job entails adjusting the GlideScore of each ligand by a function based on a similarity score. Glide can use either of two methods for calculating this similarity score: Standard Similarity Scoring and Weighted Similarity Scoring. In Standard Similarity Scoring, each ligand is compared to a set of "known actives." All atom pairs that a given ligand has in common with a given active compound are counted equally, and the similarity score for a ligand with respect to an active compound depends only on the number of such matches, normalized by the total number of atom pairs in the two molecules. For Standard Similarity Scoring, the file containing the active compounds need not be specified until the ligand docking job.

Weighted Similarity Scoring, as the name implies, counts some AP matches more than others. This scoring method requires specification of a set of "decoy" compounds (often referred to as "inactives," but not necessarily known experimentally to be inactive) in addition to the actives, and uses both sets during a grid generation job to calculate a set of calibrated weights for the atom pairs found in the active compounds. Specifically, APs found only in active compounds receive higher weights than those found in both actives and inactives. In a subsequent docking job, when Glide computes the similarity score of a given ligand to be docked, it applies these weights to the atom pairs found in that ligand. An AP found in a ligand that was not encountered during weight calibration (i.e. did not occur in either the actives or the inactives) is assigned an "average" weight of 1.0 by default. Weighted similarity scores may help identify ligands that share specific characteristics with the actives that are absent from the inactives, i.e. a common "pharmacophore."

For either method of similarity scoring, the Glide docking job adjusts the GlideScore of each ligand based on the highest similarity of that ligand to any active compound. If the

highest similarity is *s*, the adjustment is a penalty function *f(s)* added to the GlideScore computed in the usual way, with the following "ramp" form:

$$f(s) = \begin{cases} P_{max} & s < s_{min} \\ P_{max}(1 - (s - s_{min})/(s_{max} - s_{min})) & s_{min} \leq s < s_{max} \\ 0 & s \geq s_{max} \end{cases}$$

The defaults are: P_{max} = 6.0 GlideScore units, s_{min} = 0.3, s_{max} = 0.7. All three parameters are selectable in Maestro, subject only to 0.0 <= s_{min} <= s_{max} <= 1.0. In particular, P_{max} may be negative, with the effect of rewarding ligands that are dissimilar to the specified actives. This is useful for finding "new leads" that bind well to a target, but are in a different chemical family from previously known binders.

6.9.2 Similarity Folder Features: Grid Generation

When the Glide task Set up grids is selected in the Settings folder, the Generate weights option in the Similarity folder becomes available:

Generate weights

Selecting this option makes the following features available for specifying the active and inactive compounds to be used in determining weighted similarities.

Filename of known actives

Select a file containing the active compounds to be used in calibrating similarity weights. For this purpose, the file must contain at least two structures. Note that a file of actives must be specified again in the docking job, for both standard and weighted similarity. For best results in both cases, these compounds should span the range of functionality and molecule size of known actives. Including multiple compounds that are closely related may be inefficient, because any given ligand will have nearly identical similarity scores with respect to each of them. Including diverse compounds, by contrast, will help ensure that a ligand that is likely to be a good binder will have high similarity to at least one of them.

Filename of known inactives

Select a file containing the decoy compounds to be used in calibrating similarity weights. The file must contain at least one structure. Note that these need not be "known inactives" in the sense of experimental evidence that they do not bind to the target receptor. Instead, they should be compounds that lack certain features known to be characteristic of the

actives, or else randomly chosen compounds that *probably* don't bind or *probably* are dissimilar to the actives. The inactives should be a diverse set of compounds, but there are no specific criteria for selecting them.

The files of active and inactive compounds may be in either Maestro or MDL SD format.

Percentage of inactives to keep

This must be an integer from 1 to 100. Structures from the inactive compounds file are selected at random until the specified percentage (approximately) is obtained, and only the selected structures are used in weight calibration. For best results, choose the percentage based on the relative size (number of structures) in the active and inactive files, in order to

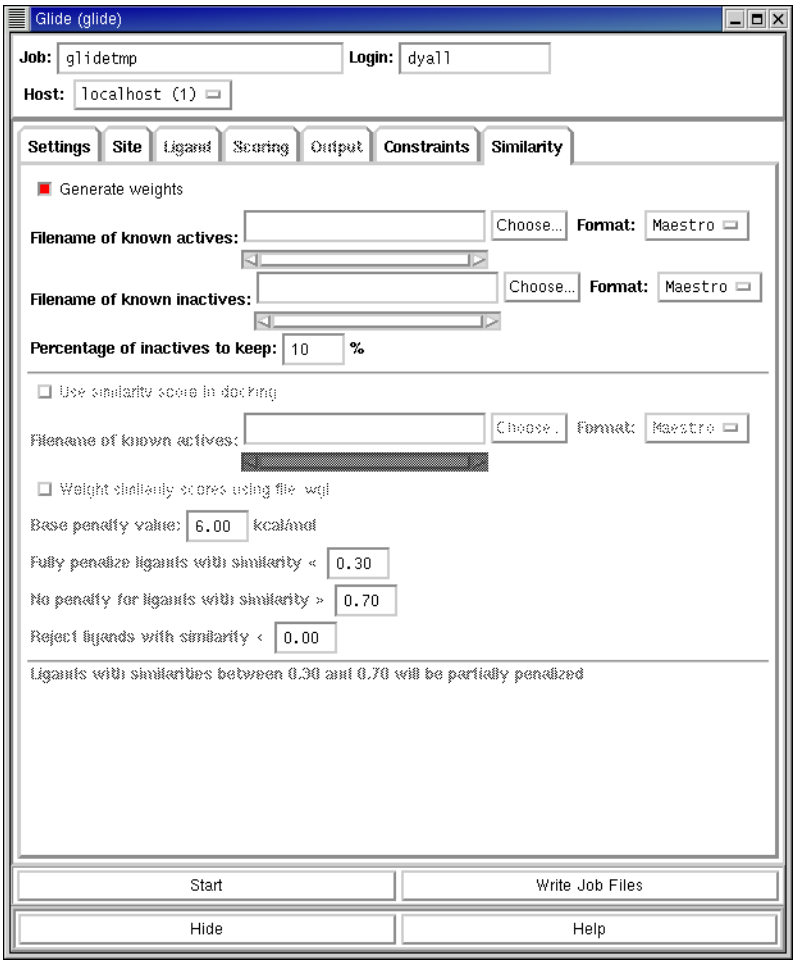


Figure 6.12. The Similarity ***folder:*** Generate weights.

obtain a total number of inactives between 5 and 15 times the number of actives. Weight calibration may produce a message stating that it did not converge (more likely the higher the ratio of inactives to actives), but this is not a problem. A valid weights file is produced in any case, and contains the "best" weights obtained with the given structures.

6.9.3 Similarity Folder Features: Docking

When the Glide task Dock one or more ligands is selected in the Settings folder, the Use similarity score in docking option becomes available in the Similarity folder:

Use similarity score in docking

Selecting this option makes the following features available:

- Filename of known actives
- Weight similarity scores using file *gridbase*`.wgt`
- Base penalty value ___ kcal/mol
- Fully penalize ligands with similarity < ___
- No penalty for ligands with similarity > ___
- Reject ligands with similarity < ___

Filename of known actives

Select a file containing the active compounds to be used in atom pair matching for computing similarity scores. For weighted similarity, this file should normally be the same one used to calibrate the weights, because weights calibrated with one set of actives may not give good results when used with a different set.

Weight similarity scores using file *gridbase*`.wgt`

Turn this toggle on to specify weighted similarity scoring. Note that the weights will be taken from a file (produced in a grid generation job run with the "Generate Weights" toggle on) with the same "base name" as the Glide docking grids, and this name is not separately adjustable. In other words, if you are using docking grids called *mybase*`.grd`, etc., and you launch your job from Maestro, you must use similarity weights from *mybase*`.wgt`. In order to specify weights from a different file (produced in a different grid generation job), you must edit your Glide input file by hand and run it from the command line rather than from Maestro.

Base penalty value

The value entered here is P_{max} in the expression above, which is added to the GlideScore of ligands that fall below the specified similarity threshold s_{min}. The default for this base penalty is 6.0, but any real value is allowed. A value of 0.0 is equivalent to running without

similarity scoring, and negative values result in rewards rather than penalties for ligands that are dissimilar to the compounds in the actives file.

Fully penalize ligands with similarity < ___

The value entered here is the similarity threshold s_{min}, below which ligands get the full penalty (or reward). This is a real number between 0.0 and 1.0, with a default of 0.3, and must be less than or equal to the "Zero Penalty" threshold in the next item. Note that it is always the maximum similarity of a ligand to any active compound that determines its GlideScore adjustment, whether that adjustment is a penalty or a reward for dissimilarity.

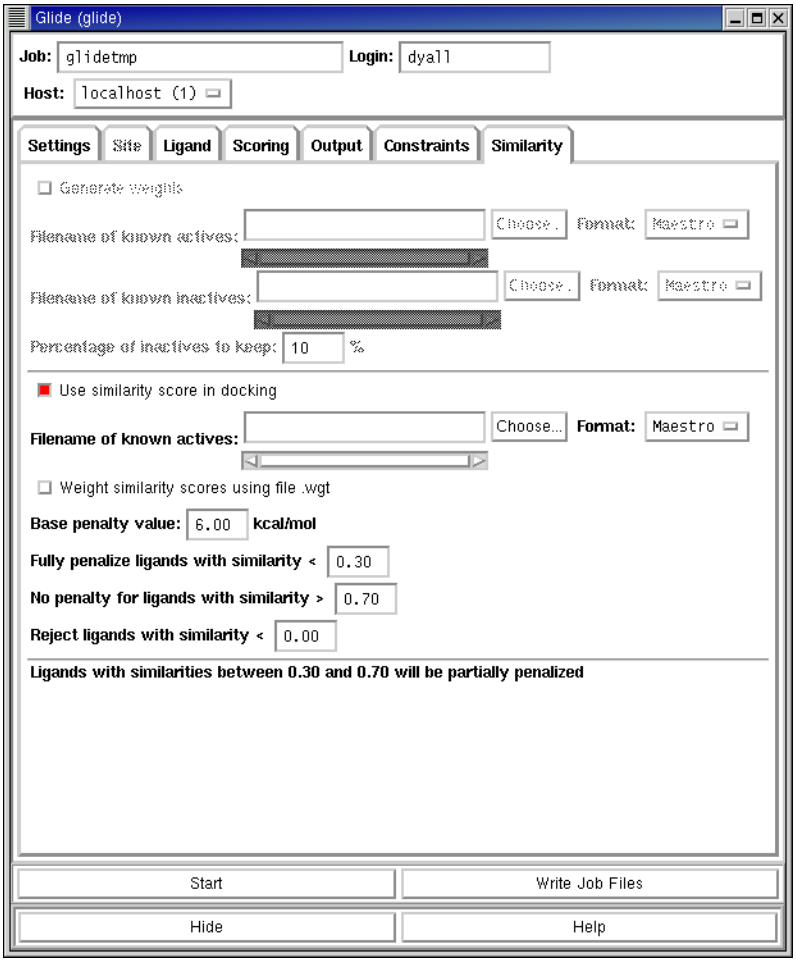


Figure 6.13. The Similarity ***folder:*** Use similarity score in docking.

No penalty for ligands with similarity > ___

The value entered here is the similarity threshold s_{max}, above which ligands get no penalty (or reward). This is a real number between 0.0 and 1.0, with a default of 0.7, and must be less than or equal to the "Full Penalty" threshold in the previous item.

Reject ligands with similarity < ___

This must be a positive number between 0.0 and 1.0. The default is 0.0. Glide will skip any ligand whose maximum similarity to any of the actives is less than this value. Thus the default value of 0.0 means do not skip any ligands, while the maximum value of 1.0 means skip any ligand that is not identical to, or a stereoisomer of, one of the known actives.

6.10 The Pose Viewer

The Pose Viewer is not part of the Glide panel, but has its own panel which you open from the Tools menu. Use the Glide Pose Viewer panel to display the contents of a Pose Viewer file written by Glide and to write out selected poses to another file.

To open the Pose Viewer, choose Glide Pose Viewer from the Tools menu.

6.10.1 Pose Viewer Panel Features

File Selection

Click Open to display a file selector. Use this selector to choose a pose file to display. Such files typically end in `_pv.mae`. The File text box is read-only—file names cannot be directly entered there.

Receptor Display

The first structure in the pose file is assumed to be the receptor. The receptor can be displayed or undisplayed by toggling the Display button. Note that this action affects the numbering of the molecules in the Workspace. In particular, the initial Pose Viewer display will have the receptor as molecule 1 (or 1 to *n*, if it is composed of *n* molecules) and the displayed ligand pose as molecule 2 (or *n*+1). When the receptor is undisplayed, the displayed ligand becomes molecule 1, and remains molecule 1 when the receptor is redisplayed. This change in numbering affects ASL commands that you might want to employ to restrict the Workspace display to the region of the receptor near the ligand, to color certain species in certain ways, or for some other purpose (see the *Maestro User Manual* or Maestro online help for information on ASL expressions). Note, however, that when the next ligand is displayed, the protein will again become molecule 1 (or 1 to *n*) while the new ligand will be molecule 2 (or *n*+1).

Glide Pose Viewer (pose)
Poses
H-Bonds
Contacts
Open...
File: /home/saunders/Documentation/FirstDiscovery_Tutorial/tutorial/glide/dock_confgen_pv.mae
Receptor: receptor@0
Display
Index Title Lig # Conf # Pose # G-Score E-Model Energy HBnd Good vdW Bad vdW Ugly vdW
1 H01 ligand 1 127 168 -9.39 -83.5 -52.5 3 284 4 2
2 H18 ligand 18 48 99 -9.35 -72.3 -43.2 1 395 24 0
3 H10 ligand 10 1 4 -9.23 -81.8 -51.0 1 338 7 1
4 H12 ligand 12 60 213 -8.96 -80.0 -48.5 1 325 9 1
5 H17 ligand 17 99 138 -8.95 -72.6 -50.5 0 398 5 0
6 H11 ligand 11 10 328 -8.90 -78.0 -49.9 1 336 5 1
7 H09 ligand 9 6 314 -8.89 -77.6 -49.7 1 328 5 1
8 H05 ligand 5 48 267 -8.77 -65.0 -39.7 1 359 16 0
9 H16 ligand 16 90 91 -8.77 -56.2 -38.2 3 337 20 2
10 H02 ligand 2 25 187 -8.65 -76.7 -48.9 1 321 6 1
11 H03 ligand 3 15 252 -8.62 -75.0 -47.8 1 301 5 1
12 H15 ligand 15 41 62 -8.44 -58.1 -35.3 3 353 19 3
Previous
Next
Visualize H-Bonds to Receptor
Visualize Contacts to Receptor
Write Displayed Poses...
Hide
Help

Figure 6.14. The* Poses *folder of the* Pose Viewer *panel.

Ligand/Pose Scroll List

This part of the panel lists all of the ligand poses reported by Glide. The entries are listed with a numerical index, a title, and various properties of the combination of the posed ligand and the receptor. The index is the ligand's position relative to other ligands in the pose file. (Note that the first indexed ligand is actually the second structure in the file since the first structure is the receptor.) The Title field shows text that typically is a descriptive name or registry identifier for the ligand. This 30-character field is carried over from the title or molecule-name record of the input Maestro or MDL SD file, or, for PDB files, is generated from the file name. The ligand number, conformation number, and pose number are also shown. The G-Score (GlideScore), Emodel, and Energy values calculated by Glide are shown whenever they are available. The Hbond, Good vdW, Bad vdW, and Ugly vdW values shown are the numbers of H-bonds and contacts between the posed ligand and the receptor, evaluated by Maestro according to the current H-bond and contact measurement criteria.

Selection of Ligand Poses

A single ligand pose can be selected and displayed by clicking on its entry row. Additional entries can be added to the selection (and to the Workspace display) by holding down either CTRL or SHIFT and clicking on their entry rows. This might be done to visualize two or more poses simultaneously or to select a subset of poses to write to an external file using the Write Displayed Poses button. Control-click acts as a toggle, and thus can be used either to add a new entry to the Workspace or to de-select an existing entry. Unlike some other selection paradigms, the Pose Viewer does not allow a range of entry rows to be added to the selection via a single shift-click.

Modifications in the Workspace

Modifications of selected Pose Viewer structures affect only the Workspace. For example, if a pose is selected in the Pose Viewer, modified in the Workspace (say, some atoms are deleted), unselected in the Pose Viewer, and then reselected, the pose will be displayed in its *original* state. That is, modifications are discarded and do not affect the original pose in the pose list.

Previous

This button selects and displays the entry row that directly precedes the first selected entry row. Click Previous at the start of the list to take you to the end.

Next

This button selects and displays the entry row that now directly succeeds the last-most selected entry row. Click Next at the end of the list to return you to the top.

Visualize H-Bonds to Receptor

This button displays markers for hydrogen bonds between receptor and non-receptor molecules in the Workspace. This button is enabled only when the receptor and at least one ligand pose are displayed in the Workspace. Clicking this button alters the H-bond Set 1, as listed in the H-Bonds folder, to be the receptor molecule (or molecules) and the H-bond Set 2 to be all other molecules. This action also displays the H-bond markers. The hydrogen bond settings can be viewed and altered in the H-Bonds folder, which can be opened either from the Glide Pose Viewer Panel or from the Measurements Panel on Maestro's main menu bar.

Visualize Contacts to Receptor

This button displays the bad and ugly ("really bad") contacts between the receptor and non-receptor (i.e., ligand) molecules in the Workspace. This button is enabled only when the receptor and at least one other pose structure are displayed in the Workspace. Clicking this button alters Contact Set 1, as listed in the Contacts folder, to be the receptor molecule (or molecules) and Contact Set 2 to be all other molecules. This action also displays the contact markers. The contact settings can be viewed and altered using controls in the Contacts folder, which can be opened either from the Glide Pose Viewer Panel or from the Measurements panel on Maestro's main menu bar.

Write Displayed Poses

This button is enabled only when the receptor, or at least one of the ligand pose structures, is displayed in the Workspace. This button opens the PoseWrite panel, which enables displayed pose structures to be written out to another file. The operation of the PoseWrite panel is described in Section 6.10.2.

H-Bonds Folder

This folder contains controls that set the criteria for display of hydrogen bonds. The H-Bond controls in this folder operate in the same way as those in the H-Bonds folder in the Maestro Measurements panel. See the *Maestro User Manual* or the Maestro online help for more information.

Contacts Folder

The controls in this folder set the criteria for display of contacts. Contacts can be good, bad, or ugly. Any combination can be shown (or hidden) by using the Mark Contacts buttons. Good contacts are displayed as green, bad contacts as orange, and ugly contacts as red. The controls in this folder operate in the same way as those in the Contacts folder in the Measurements panel. See the *Maestro User Manual* or the Maestro online help for more information.

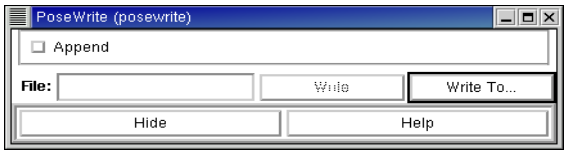


Figure 6.15. The** PoseWrite **panel.

6.10.2 The PoseWrite Panel

The PoseWrite panel allows you to write out to a Maestro file the pose file structures selected for display in the Glide Pose Viewer panel.

To open the PoseWrite panel, click Write Displayed Poses on the Glide Pose Viewer panel.

To save a file in the current directory, enter the name in the File text box and press the ENTER key or click the Write button. The file will be written using the current settings (see Append, below). Maestro accepts absolute (e.g., `/home/joe/file1.mae`) or relative (e.g., `file1.mae`) file names.

If you are unsure which directory to use, or if you are concerned about overwriting an existing file, click the Write To button. The Write Pose File file-selector panel is displayed. When you locate the directory and determine that your file name is unique, type the name in the Selection text box and press ENTER or click OK. For more information on using a file selector, click the Help button on the Write Pose File panel.

Append

When the Append button is enabled, the poses currently selected on the Glide Pose Viewer panel are appended to the file whose name appears in the File text box. When the button is deselected, the file overwrites any existing file of the same name.

6.11 Glide Utilities

FirstDiscovery provides command-line utilities and applications to aid in structure preparation, structure file format conversion, and structure database handling. The Glide module has its own set of utilities. The command-line utilities `glide_sort`, `glide_rescore`, and `para_glide` are summarized in this section.

6.11.1 glide_sort

Purpose: Re-ranks Glide poses by custom criteria or combines job outputs into one file.

Syntax:

```
$SCHRODINGER/utilities/glide_sort mode [options] Glide-pose-files
```

Modes of Operation:

At least one of these options is required:

`-o` *output-file*	Write the best-scoring poses to *output-file.*
`-r` *report-file*	Create a report of the best scores in *report-file.*
`-R`	Write a report of the best scores to standard output.

Sorting Options:

`-use_dscore`	Default. Sort poses based on the "docking score" in Glide output poses."Docking score" is a placeholder for the property on which you would like to sort poses. Initially it is equal to GlideScore, but `glide_rescore` can be used to replace it with other values (see Section 6.11.2.)
`-use_gscore`	Sort poses based on GlideScore. Overrides use of "docking score" field.
`-use_cvdw`	Sort poses based on Coulomb-van der Waals energy, E(CvdW). Overrides use of "docking score" field.
`-use_emodel`	Sort poses based on model energy score, Emodel. Overrides use of "docking score" field.
`-nosort`	Don't sort the poses.

Output Options:

`-n` *nreport*	Retain only the *nreport* lowest-scoring poses.
`-norecep`	Don't include the receptor structure in output files.
`-best`	Keep only the single best pose for each ligand in each input pose file.
`-best-by-lignum`	Keep only the single best pose for each ligand with a given *lignum* value.
`-best-by-title`	Keep only the single best pose for each ligand with a given title.
`-h`	Print help message and quit.
`-v`	Print version number and quit.

Custom Scoring Function Options:

Setting any of these custom scoring function terms causes `glide_sort` to sort only on this custom function instead of the standard `-use_` sorting options.

`-gscore` *coef*	GlideScore coefficient (default = 0.0) for custom scoring function.
`-cvdw` *coef*	E(CvdW) coefficient (default = 0.0) for custom scoring function.
`-internal` *coef*	E(internal) coefficient (default = 0.0) for custom scoring function.
`-emodel` *coef*	Emodel coefficient (default = 0.0) for custom scoring function.
`-offset` *coef*	Energy offset (default = 0.0) for custom scoring function.

Filter Options:

`-hbond_cut` *cutoff*	Filter cutoff for H-bond energy (default = 0.0).
`-cvdw_cut` *cutoff*	Filter cutoff for E(Cvdw) (default = 0.0).
`-metal_cut` *cutoff*	Filter cutoff for metal-ligation energy (default = 0.0).
`-emodel_cut` *cutoff*	Filter cutoff for Emodel (not used by default).
`-nofilter`	Don't use filter cutoffs at all.

6.11.2 `glide_rescore`

Purpose: Replaces the "docking score" properties in Glide pose output files with different values, so that the `glide_sort` "best-by-title" option can be used to combine different screens. See Section 6.11.1.

Syntax:

`$SCHRODINGER/utilities/glide_rescore [`*options*`]` *pv-or-lib-files*

Options:

`-rank`	Replace "docking_score" with ligand rank (default mode).
`-offset` *value*	Replace "docking_score" with GlideScore plus this offset.
`-average`	Calculate the average GlideScore over all the poses; no output besides this average is produced.
`-top` *number*	Average only the top *number* poses.
`-every` *number*	Print running averages every multiple of *number* poses.
`-o` *output-file*	Output to this file name, instead of default name (*input-file*`.rank.mae` or *input-file*`.offset.mae`).

`-h`	Print help message and quit.
`-v`	Print version number and quit.

Only one of `-rank`, `-offset`, and `-average` can be used at a time. If none is specified, `-rank` is assumed.

Using `-top` and/or `-every` implies `-average` mode.

6.11.3 `para_glide`

Purpose: Submits batches of ligand structures to multiple processors for Glide docking jobs.

Syntax:

```
$SCHRODINGER/utilities/para_glide -i inp-file [options]
```

Options:

`-n` *njobs*	Number of subjobs to prepare.
`-f` *firstlig*	First ligand to include.
`-l` *lastlig*	Last ligand to include.
`-j` *jobnum*	Subjob number to prepare.
`-x`	Launch jobs after writing input files.
`-s`	Split input ligand file by subjob.
`-o`	Have log file written directly to output directory (only meaningful if `-x` option also used).
`-h`	Print help message and quit.
`-v`	Print version number and quit.

Split the Glide job specified in the *inp-file* into smaller subjobs for distributed execution. The ligands between *firstlig* and *lastlig* (inclusive) are separated into *njobs* equal-sized batches. If omitted, *njobs* defaults to 1, *firstlig* defaults to 1 and *lastlig* defaults to 0, which is interpreted to mean the final ligand in the ligands file.

Each use of `para_glide` creates two scripts: *job*`_report.sh` and *job*`_status.sh`. The *job*`_report.sh` script collects the output (poses) from subjobs created by `para_glide`, and produces a single pose file and a single report file that summarizes the best poses in the entire job suite. The *job*`_status` script can be run while the job suite is running or afterward. It summarizes the disposition of each job: whether it finished normally, died, was terminated, stopped, and so on, using the conventional Schrödinger job control terms. (See the *Maestro User Manual* for information about job monitoring and job control.)

The `-j` option is useful for preparing only a single subjob. A value of 0 is equivalent to not specifying `-j` at all, and all subjobs will be printed. Negative values are not permitted.

By default, all jobs use the same input ligand file, each job reading out of it just the ligands pertinent to that job. If the `-s` option is given, a new ligand file is written for each subjob, containing only the ligands for that subjob. This is likely to be more efficient if there are a large number of ligands.

If the `-x` option is given, then the jobs are launched after the input files are written. Any additional arguments you give on the command line are passed on to the `impact` command. Therefore, you can have the jobs run on a remote machine by specifying `-HOST` *hostname*.

Chapter 7: Liaison

7.1 Brief Description of Liaison

A Liaison simulation combines a molecular-mechanics calculation with experimental data to build a model scoring function used to correlate or to predict ligand-protein binding free energies. A method of this type is called a Linear Response Method (LRM), a Linear Interaction Approximation (LIA), or a Linear Interaction Energy (LIE) method.

A novel feature of Liaison is that the simulation takes place in implicit (continuum) rather than explicit solvent—hence the name Liaison, for Linear Interaction Approximation in Implicit SOlvatioN. The explicit-solvent version of the methodology was first suggested by Aqvist (Hansson, T.; Aqvist, *J. Protein Eng.* **1995,** *8,* 1137-1145), based on approximating the charging integral in the free-energy-perturbation formula with a mean-value approach, in which the integral is represented as half the sum of the values at the endpoints, namely the free and bound states of the ligand. The empirical relationship used by Liaison is shown below:

$$\Delta G = \alpha\,(\langle U^{b}_{vdw}\rangle - \langle U^{f}_{vdw}\rangle) + \beta\,(\langle U^{b}_{elec}\rangle - \langle U^{f}_{elec}\rangle) + \gamma\,(\langle U^{b}_{cav}\rangle - \langle U^{f}_{cav}\rangle)$$

Here < > represents the ensemble average, *b* represents the bound form of the ligand, *f* represents the free form of the ligand, and α, β, and γ are the coefficients. U_{vdw}, U_{elec}, and U_{cav} are the van der Waals, electrostatic, and cavity energy terms in the Surface Generalized Born (SGB) continuum solvent model. The cavity energy term, U_{cav}, is proportional to the exposed surface area of the ligand. Thus, the difference:

$$\langle U^{b}_{cav}\rangle - \langle U^{f}_{cav}\rangle$$

measures the surface area lost by contact with the receptor. The net electrostatic interaction-energy in continuum solvent is given by:

$$U_{elec} = U_{coul} + 2\,U_{rxnf}$$

where U_{coul} is the Coulomb interaction energy and U_{rxnf} is the SGB-solvent reaction-field energy. (The factor of 2 compensates for the division by 2 made in the definition of the reaction-field free energy.)

In most applications, the coefficients α, β, and γ are determined empirically by fitting to the experimentally determined free energies of binding for a training set of ligands. In such applications, Liaison's Simulate task is used to calculate the values of U_{vdw}, U_{elec}, and U_{cav} for the bound (complexed) and unbound (free) states of the training-set ligands, and its Analysis task is used to derive values for the α, β, and γ fitting coefficients. The fitted equation can then be used to predict the binding affinities of additional ligands.

Liaison also calculates the GlideScore over the course of the LRM simulation. The average GlideScore can then be used to predict binding energies using the alternate model:

$$\Delta G = a(<GlideScore>) + b$$

where a is the GlideScore coefficient (Slope) and b is a constant (Intercept).

The GlideScore alternative binding energy model can be selected in the Liaison panel's Analysis folder. The analysis task will then derive values for the a and b fitting coefficients. The GlideScore binding energy model is discussed further in Section 7.3 on page 118 and Section 7.9 on page 132.

7.2 Liaison Simulations

When you start a Liaison simulation from the Maestro interface (or when writing job files for a manual submission), several actions occur in the background. This section describes some of these background features.

7.2.1 Liaison Directory Structure

Figure 7.1 on page 115 shows a schematic overview of the Liaison directory structure, where "Maestro Working Directory" is the directory in which Maestro is running when you start a Liaison job or elect to Write Job Files. Files in brackets [ ] are created only with Liaison Dynamics and HMC jobs.

Note: When you run a Liaison job remotely, make sure the job files are accessible from the remote machine.

7.2.2 Directories Created

Master Liaison Run Directory

This directory is created by Maestro under the Maestro working directory on starting the job or writing job files. The Master Liaison Run Directory name is defined by the text entered in the Job text box in the top left corner of the Liaison panel. Hence, *jobname* = Master Liaison Run Directory. (Your choice for the Job text string is denoted as *jobname*). The Master Liaison Run Directory specifies a relative rather than an absolute path.

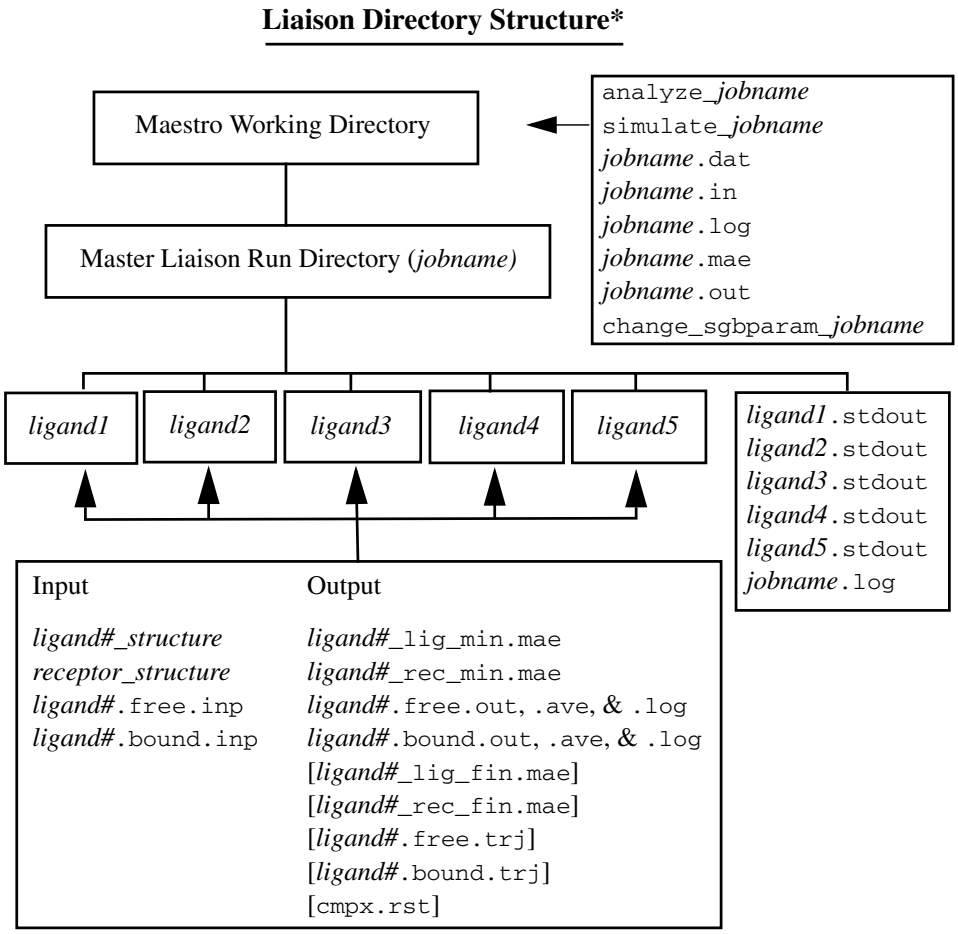


Figure 7.1. The Liaison directory structure.

* Files in brackets [ ] are created only with Liaison Dynamics and HMC jobs.

Ligand Directory (or Directories)

The directory or directories created under the Master Liaison Run Directory. The names of the directories are defined by ligand names that you specify.

7.2.3 Files Created

In the Maestro working directory, the following files are created:

- `change_sgbparam_`*jobname*: Utility to modify the SGB solvation parameters for the input files in all the ligand subdirectories.
- `simulate_`*jobname*: The main script Maestro uses to dispatch the Liaison simulation.
- *jobname*`.mae`: The receptor (or receptor/ligand) structure file in Maestro format. This file is written by the Maestro interface.
- `analyze_`*jobname*: Script to run a Liaison analysis (fitting or predicting) job.
- *jobname*`.in`: Input file for a Liaison analysis job.
- *jobname*`.dat`: Data file for a Liaison analysis job.
- *jobname*`.log`: Log file for a Liaison analysis job.
- *jobname*`.out`: Output from a Liaison analysis job.

In the Master Liaison Run Directory, the following files are created:

- *jobname*`.log`: Log of ligands submitted.
- For each ligand in the job, two files are created, but these are generally of interest only if an error has occurred:
  - *ligand#*`.log`
  - *ligand#*`.stdout`

In the Ligand directory or directories, the following files are created:

- *jobname*`.free.inp`: The Liaison input file for simulation of the free ligand.
- *jobname*`.bound.inp`: The Liaison input file for simulation of the ligand-receptor complex.
- Link to the receptor structure file *jobname*`.mae` that Maestro wrote in the Maestro working directory. In a single ligand job, the ligand is also contained in the structure file if it was taken from the Workspace.
- Link to the location of the ligand file, which may or may not be the same as the receptor structure file.
- *jobname*`_lig_min.mae`: Final minimization structure for the ligand from the bound simulation, when minimization is used as the sampling method or when the ligand is minimized prior to Hybrid Monte Carlo or Molecular Dynamics sampling.
- *jobname*`_rec_min.mae`: Final minimization structure for the receptor from the

bound simulation, when minimization is used as the sampling method or when the complex is minimized prior to HMC or MD sampling.

- *jobname*_lig_fin.mae: Final simulation structure for the ligand from the bound simulation when HMC or MD sampling is used.
- *jobname*_rec_fin.mae: Final simulation structure for the receptor from the bound simulation, when HMC or MD sampling is used.
- Other output files from the simulations, including energy output (*.out, *.log, *.ave) and trajectory files from sampling (*.trj).

An example Liaison directory structure for a Multiple Ligand/Single Receptor job named 1bkm is shown below:

```
-----------------------------------------------
% pwd
/home/user
% ls
1bkm/
simulate_1bkm
1bkm.mae
% ls 1bkm
1bkm_3m_1/
1bkm_3m_2/
1bkm_3m_3/
-----------------------------------------------
```

In this example, 1bkm represents the character string entered into the Job text box, and thus 1bkm/ is the name of the Master Liaison Run Directory. The underlying ligand directories are 1bkm_3m_1/, 1bkm_3m_2/, and 1bkm_3m_3/.

7.2.4 Liaison Simulation Requirements

The Liaison simulation requirements are:

- A receptor structure (read into Maestro before starting the job)
- One or more ligands

If there is one ligand, it can either be taken from the Workspace display or read from a file.

For two or more ligands, a text file containing ligand names you define and the full path to the location of the ligand files on disk are required. This text file can reside in any directory. The format is:

```
LigandName [space] LigandLocation
```

Example:

```
Ligand1   /home/username/LIA/H01_lig.mae
Ligand2   /home/username/LIA/H02_lig.mae
Ligand3   /home/username/LIA/H03_lig.mae
Ligand4   /home/username/LIA/H04_lig.mae
Ligand5   /home/username/LIA/H05_lig.mae
```

The sampling methods available are:

- Energy Minimization
- Hybrid Monte Carlo (HMC)
- Molecular Dynamics (MD)

By far the fastest method is energy minimization. Even though this method gives only a snapshot of the possible ligand-receptor configurations, studies to date have shown that it gives predicted binding affinities that are reasonably close to those obtained with HMC or MD. Minimization is thus an attractive choice when large numbers of ligands are to be studied. Those predicted to be most active might then be re-examined using a simulation protocol (HMC is recommended), if desired.

7.3 Liaison Analysis

A Liaison analysis calculation uses the results of completed Liaison simulations to fit the binding energy model coefficients to the binding energies of known ligands or to predict the binding energies of new ligands. To use the results of earlier simulations, a Liaison Analysis job must be run from the same directory (Maestro working directory), and the same text string must be entered into the Job text box (to identify the Master Liaison Run Directory). In addition, the supplied ligand names must be the same as those used in the simulation calculations.

7.3.1 Fitting the Simulation Results to Experimental Data

A Liaison fitting calculation requires a text file containing a list of the ligand names and their associated experimental binding energies. This text file can reside in any directory. The format is:

```
LigandName [space] BindingEnergy
```

Example:

```
Ligand1   -7.32
Ligand2   -7.73
Ligand3   -9.20
```

```
Ligand4    -8.06
Ligand5    -10.01
```

The ligand names must be the same as those used in the preceding Liaison simulation calculations.

Units of kcal/mol are assumed. If other units are employed (kJ/mol, pK_i, or IC_{50}), the numerical results are correct in those units, but "kcal/mol" is still printed.

Before starting the fitting calculation, make sure that you are in the Maestro working directory and enter the name of the Master Liaison Run Directory in the Job text box in the upper left corner of the Liaison panel.

7.3.2 Predicting Binding Affinities of New Ligands

To run a Liaison prediction calculation, the following items are required:

- Values for the coefficients of the selected binding energy model equation (usually taken from a Liaison fitting calculation).
- A list of ligand names, either contained in a file or entered as a comma- or space-separated list in Maestro. If the ligand names are in a file, you can separate them by one or more spaces, a comma, or a carriage return.

7.4 The Liaison Panel

To run a multiple-ligand job or to fit calculated Liaison results to known binding energies, you must supply an external file:

- Multiple-ligand jobs require an input file listing the ligands to be used and the locations of the associated structure files.
- Fitting Liaison results requires a file that lists the ligands and their binding energies.

When you submit a Liaison job or click the Write Job Files button, Maestro writes several files and directories to disk in the Maestro working directory (the directory Maestro is in when the job is started or job files are written). Specifically, it uses the name entered in the Job text box of the Liaison panel to create a Master Liaison Run Directory under the launch directory, if the directory does not already exist. Maestro then also creates one or more ligand directories under the Master Liaison Run Directory.

To open the Liaison panel, choose Liaison from the Applications menu in the Maestro main menu bar.

The Liaison panel has five tabbed folders:

- Settings
- System
- Parameters
- Constraints
- Analysis

7.5 The Settings Folder

The Job Type option menu at the top of the folder selects Simulation or Analysis as the basic function of the Liaison task. This folder also specifies the number of processors to use (subject to the number of Liaison licenses) in concurrent Liaison simulations on a multi-processor machine. A Simulation job requires additional settings in the System, Parameters, and Constraints folders. It is usually convenient to proceed through the folders in this order, but any order is allowed.

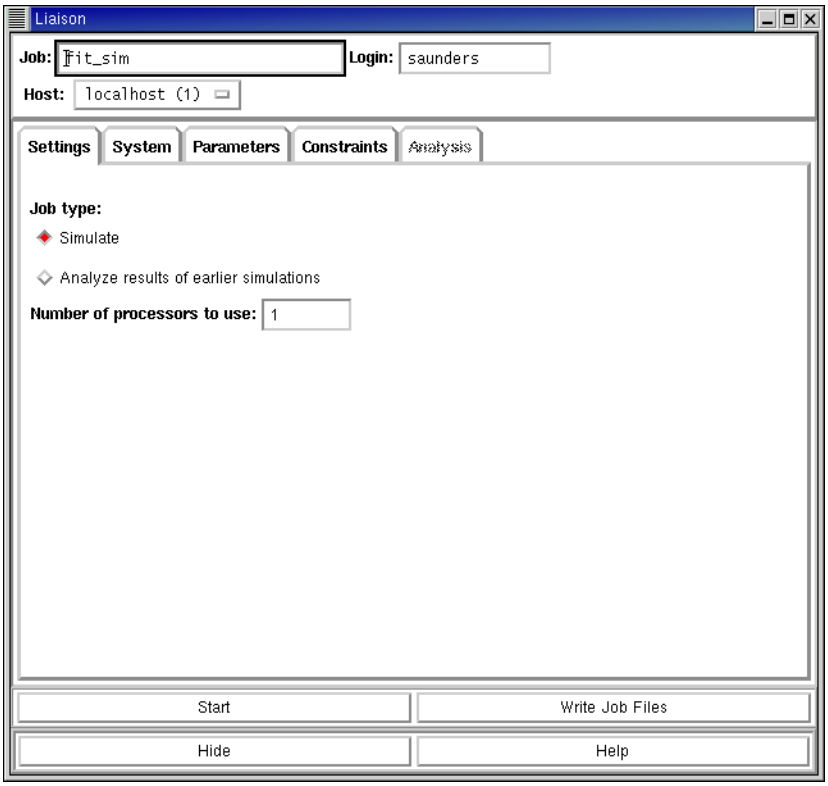


Figure 7.2. The* Settings *folder of the* Liaison *panel.

The selection options in the Settings folder are:

- Job type (buttons)
- Number of processors to use (text box)

This section also provides instructions for killing Liaison jobs from the command line.

Job type

This pair of buttons determines the type of Liaison job to be run.

Simulate

Use this selection to set up a Liaison simulation. It allows access to all Liaison folders (System, Parameters, and Constraints) except for the Analysis folder.

Analyze results of earlier simulations

Use this selection to analyze the output of completed Liaison simulations by fitting calculated results to empirical binding energy values or by predicting the binding energy of new ligands. This selection deactivates the System, Parameters, and Constraints folders and activates the Analysis folder.

Number of processors to use

Liaison simulation jobs can take advantage of multiple processors to perform distributed processing. Use this text box to specify the number of processors on which to run concurrent Liaison simulations. For example, if there are 10 ligand/receptor combinations (for a total of 20 jobs — 10 “free” and 10 “bound”) and there are 8 processors, setting this number to 4 will launch 4 jobs when you click the Start button. On each processor, when one job completes, another job will start, until all simulations have been submitted.

Note: this option indicates how many processors on the same machine to run simultaneously; it is meaningless when jobs are submitted to a batch queue, where each of the ligands is independently queued.

This option is not available when the Job type is Analyze results of earlier simulations, as analysis jobs are much faster than simulation jobs.

7.6 The System Folder

This folder sets the type of simulation to be run and defines the system and the source of the ligand or ligands to be used. The key option is Simulation type, and the options are:

- Multiple ligands, single receptor
- Single ligand, single receptor

Each option affects the selection options offered in the remainder of the panel. These options are described below, together with the selection options that pertain to each. Selecting certain options dims other options that do not apply to the selected options.

For a Single ligand, single receptor simulation, you have two options:

- Both receptor and ligand are loaded to the Workspace. The structures do not have to be loaded in any particular order.
- Only the receptor is loaded into the Workspace. The ligand is defined via an external file.

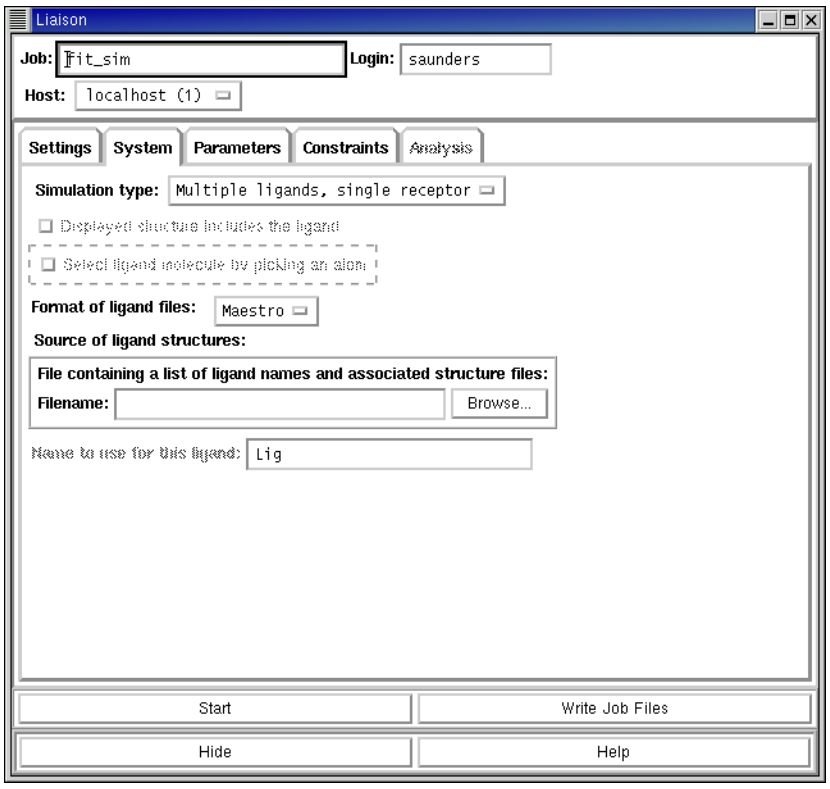


Figure 7.3. The System ***folder of the*** Liaison ***panel.***

For Multiple ligands, single receptor simulations, only the receptor can be loaded to the Workspace. The ligands are defined by specifying a file that contains the ligand names and the locations of their structure files. In each case, selecting certain options dims other control settings that do not apply to the options selected.

If the Analysis folder is active, you must change the Job Type selection to Simulate in the Settings folder before opening the System folder.

7.6.1 Multiple Ligands, Single Receptor

A multiple-ligand simulation requires one receptor structure and multiple ligand structures. The receptor structure is taken from the Workspace, which must contain only this structure. The control settings used to specify the ligands are described below.

Format of ligand files

Liaison supports the three file formats listed below for reading structures from local or network-mounted disks. Liaison does not support Mol2 formatted files, nor will it accept structures with lone pairs; remove them using the Hydrogen Treatment panel (select Hydrogen Treatment from the Edit menu) before submitting the Liaison job.

- Maestro: Maestro-written files (extensions `.mae`, `.out`, or `.dat`)
- MDL SD: SD-formatted files (extensions `.mol` for single structure files and `.sdf` for multiple structure files)
- PDB: Rutgers Center for Structural Biology Protein Data Bank files (extensions `.pdb` or `.ent`)

File containing a list of ligand names and associated structure files

The identity of a user-created text file containing the ligand names and structure-file locations can be entered directly into the Filename text box. Alternatively, the Browse button adjacent to the text box can be used to activate the Open File panel to aid in locating the file. The identity of the selected file is then displayed in the Filename text box.

The name of the text file cannot be the same as the name of the Master Liaison Run Directory (the name that is specified in the Job text box). If it is, Maestro will display an error message. The content and structure of the file is illustrated below.

```
1bkm_3m_1          /home/user/structs/1bkm_3m_1.mae
1bkm_3m_2          /home/user/structs/1bkm_3m_2.mae
1bkm_3m_3          /home/user/structs/1bkm_3m_3.mae
```

The first column contains the user-defined ligand name, and the second gives the directory path and file name of the ligand structure file. Liaison uses each ligand name to create a

correspondingly named ligand directory under the Master Liaison Run Directory. Only spaces (not tabs or commas) can separate the ligand name and file location.

7.6.2 Single Ligand, Single Receptor

A single-ligand simulation requires a receptor structure and a ligand structure. The receptor structure is taken from the Workspace. It may be alone in the Workspace or may be accompanied by the ligand to be simulated. The selection options for this Job Type are described below.

Displayed structure includes the ligand

Select this option if the ligand to be simulated is included with the receptor structure in the Workspace. Maestro then activates the Select Ligand By Picking an Atom button and expects you to pick an atom to identify the ligand. This will be the only ligand simulated.

Select ligand molecule by picking an atom

This button is active when Displayed structure includes the ligand is selected. To select the ligand, click on an atom of the Workspace structure. This visually marks the selected structure using a blue Ball & Stick representation, and everything else is taken to be the receptor. *The ligand must be marked in this manner or Maestro will consider it as part of the receptor, and will be unable to minimize or simulate any ligand in the space already occupied by this ligand.*

Format of ligand file

The file format needs to be defined when the ligand is not being taken from the Workspace, i.e., when Displayed structure includes the ligand is not selected. Liaison supports the three file formats listed below for reading structures from local or network-mounted disks. Liaison does not support Mol2 formatted files and will not accept structures with explicit lone pairs. Remove them via the Hydrogen Treatment menu before submitting the Liaison job.

- Maestro: Maestro-written files (extensions `.mae`, `.out`, or `.dat`)
- MDL SD: SD-formatted files (extensions `.mol` for single structure files and `.sdf` for multiple structure files)
- PDB: Rutgers Center for Structural Biology Protein Data Bank files (extensions `.pdb` or `.ent`)

File containing a single ligand structure

When the ligand is not in the Workspace (i.e., when Displayed structure includes the ligand is not selected), the location of the ligand structure file can be entered into the Filename

text box. Alternatively, the Browse button adjacent to the text box can be used to activate the Open File panel to aid in locating the file; the identity of the selected file is then displayed in the Filename text box.

Name to use for this ligand

This text box specifies a user-assigned name for the ligand. The default is `Lig`. The name is used to create a ligand directory under the Master Liaison Run Directory.

7.7 The Parameters Folder

The upper portion of this folder sets the Sampling method and selects options that apply to both "free" (Ligand) and "bound" (Ligand/Receptor) simulations. The lower portion contains sub-tabs with identical features that independently control Ligand Simulation and Ligand/Receptor Simulation.

If the Analysis folder is active, you must change the Job Type setting to Simulate in the Settings folder before opening the Parameters folder.

7.7.1 Sampling Method

This section includes seven options, two of which are dimmed when Minimization is chosen as the Sampling method.

Sampling method

The supported sampling methods are:

- Minimization.
- Hybrid Monte Carlo. This method employs the Hybrid Monte Carlo algorithm to sample the binding of the ligand to the receptor (or the conformation of the free ligand). See Chapter 11 for details on the Hybrid Monte Carlo task and settings. When simulation (as opposed to minimization) is used for sampling, HMC is the recommended option.
- Molecular Dynamics. This method employs a Molecular Dynamics algorithm to sample the binding of the ligand to the receptor (or the conformation of the free ligand). See Chapter 10 for details on the Dynamics task and settings.

Minimization algorithm. This menu is available for all three sampling methods (there is an option to Minimize before simulation for the HMC and MD sampling methods.)You may choose the Truncated Newton, Conjugate gradient, or Steepest descent method to locally minimize the ligand/receptor (or free ligand) geometry. See Chapter 9 for details on the Minimization task and settings.

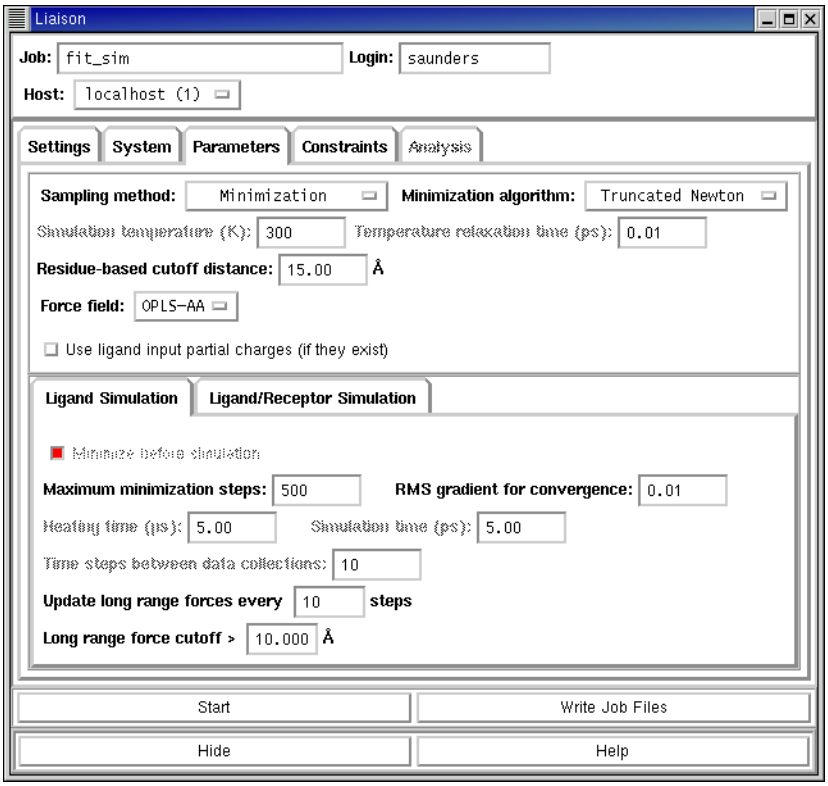


Figure 7.4. The Parameters ***folder of the*** Liaison ***panel.***

Simulation temperature

This option is available for Hybrid Monte Carlo and Molecular Dynamics sampling. It sets the simulation target temperature in Kelvin.

Temperature relaxation time

This option is available for Hybrid Monte Carlo and Molecular Dynamics sampling. It sets the time scale, in picoseconds, on which heat exchanges with the heat bath.

Residue-based cutoff distance

This text box sets the value for the cutoff distance. All pairwise interactions of an atom in residue *i* with an atom in residue *j* are included on the nonbonded pair list if any such pair of atoms is separated by this distance or less. The default value is 15 Å.

Use ligand input partial charges (if they exist)

Selecting this check box indicates that the partial charges in the input ligand Maestro files should be used instead of charges assigned by the force field atomtyper. If you have high-quality partial charges from, for example, ab initio electrostatic potential fitting, then this option can be useful.

7.7.2 Ligand Simulation and Ligand/Receptor Simulation

Depending on the sampling method chosen, some or all of the following options are available in the lower section of the folder. The two sub-tabs affect simulations on the free ligand (Ligand Simulation) and bound complex (Ligand/Receptor Simulation), respectively.

The selection options are:

Minimize before simulation

This toggle is active only for Hybrid Monte Carlo and Molecular Dynamics sampling. It places a minimization task (MINIMIZE) in the Liaison input file in front of the LRM simulation task. Its purpose is to ensure that the simulated structure does not have significant excess potential energy from bad internal contacts.

Maximum minimization steps

This text box sets the maximum number of minimization steps. This option is active only when Minimize before simulation is selected.

RMS grad for convergence

This text box sets the criterion on the rms gradient for convergence of the minimization (kcal/mol/Å). This option is active only when Minimize before simulation is selected.

Heating time

In an HMC or MD simulation, this text box sets the time (ps) over which the system is heated before the LRM task is launched to obtain averages for the van der Waals, Coulombic, reaction field, and cavity terms. The default value for the heating time is 5 ps.

When the Liaison input file is written, the heating time is converted to the value of `mxcyc` (HMC) or `nstep` (MD) written in the HMC or Dynamics sections. The table below shows how the conversion is made.

Task	Conversion Formula
HMC	Heating Time = `mxcyc * nmdmc * delt *` 6
MD	Heating Time = `nstep * delt *` 6
Min	N/A (Liaison-Minimization jobs do not do heating at all)

mxcyc = # of HMC cycles
nmdmc = # of MD steps per MC cycle
nstep = # of MD steps
delt = time step (in ps). delt is 0.002 ps for HMC, and 0.001 ps for MD.

The default 0.002 ps time step for HMC and 0.001 ps time step used for Liaison dynamics (MD) jobs cannot be modified in Maestro, but can be edited by hand in the input files. However, this is not recommended. The factor of 6 comes from the fact that heating is broken up into six equal stages.

Example: HMC method, Heating time = 15 ps. Resultant input file (HMC section only):

```
-----------------------------------------------
HMC
input cntl mxcyc 83 nmdmc 5 delt 0.002 relax 0.01 nprnt 100 seed 101
-----------------------------------------------
```

The number of HMC steps (mxcyc) is five times smaller than the number of MD steps (nstep) because each composite HMC step includes five MD steps (set by nmdmc 5 in the example above). Note also that the calculated number of steps (83) corresponds to one-sixth of the requested heating time. This is because Liaison heats the ligand-protein complex (but not the free ligand) in six equal temperature increments, each of which receives one-sixth of the total heating time. For example, for a target temperature of 300 K, the heating is done in 50 K increments of 0 – 50 K, 50 – 100 K, ... 250 – 300 K.

Given that the Liaison panel has default heating time of 5.00 ps, time step of 0.002 ps, and number of MD steps per MC step of 5, then:

mxcyc = (5.000 ps / (0.002 ps * 5) = 500. 500 divided by 6 increments is 83 (rounded) steps per increment.

Simulation time

In an HMC or MD simulation, this text box sets the simulation time for the LRM task used to determine the averages for the van der Waals, Coulombic, reaction field, and cavity terms. The default value for the simulation time is 5 ps.

When energy minimization is used for sampling, no heating is done, but a short pro-forma HMC simulation (mxcyc = 10) is carried out at 10 K to obtain the needed “averages” for the Liaison interaction quantities.

When the Liaison input file is written, the simulation time is converted to the value of mxcyc (HMC) or nstep (MD) that is written to the LRM task. The table below shows how the conversion is made:

Task	Conversion Formula
HMC	Simulation Time = `mxcyc * nmdmc * delt`
MD	Simulation Time = `nstep * delt`
Min	N/A

`mxcyc` = # of HMC cycles
`nmdmc` = # of MD steps per MC cycle
`nstep` = # of MD steps
`delt` = time step (in ps). `delt` is 0.002 ps for HMC and 0.001 ps for MD.

The default 0.002 ps time step for HMC and 0.001 ps time step used for Liaison dynamics (MD) jobs are not modifiable inside Maestro, but can be edited by hand in the input files. However, this is not recommended.

When setting up Liaison jobs from Maestro, `delt` and `nmdmc` remain constant, while `mxcyc` and `nstep` are increased/decreased to accommodate the user-specified simulation time.

Example: Dynamics method, Simulation time = 5 ps. Resultant input file (LRM section only):

```
-----------------------------------------------
sample DYNAMICS
input cntl nstep 5000** delt 0.001 relax 0.01 nprnt 100 seed 101
-----------------------------------------------
```

** 5/0.001 = 5,000 steps

The five-fold reduction in the number of HMC steps (`mxcyc`) reflects the fact that each composite HMC step includes five MD steps.

Time steps between data collections

This text box sets the number of time steps between data collections of the ensemble averages during the Liaison sampling. Entering 10 in this box (the default value) produces a line like the following in the LRM task of the Liaison input file:

```
input cntl average every 10 file lia_free.ave
```

Update long range forces every *n* steps.

This option is available only for the Truncated Newton algorithm. The default is to update long range forces every 10 steps. Between updates, estimates of these forces are used. Smaller values of *n* (more frequent updates) can be used to improve convergence, but will make the optimization slower. The maximum recommended value is 20.

Long range force cutoff > *n* Angstroms.

This option is available only for the Truncated Newton algorithm, and specifies the distance beyond which forces will be treated as "long range"—that is, updated every *n* steps, as specified in the previous option, and estimated between updates.

7.8 The Constraints Folder

The Constraints folder includes selection options that allow regions of the receptor to be defined as frozen or buffered. This folder is similar to the Constraints folder in the Impact Energy Minimization panel. Control settings that do not apply to Liaison simulations are omitted.

If the Analysis folder is open, you must first go back to the Settings folder and select Simulate for the Job type.

The selection options are described below.

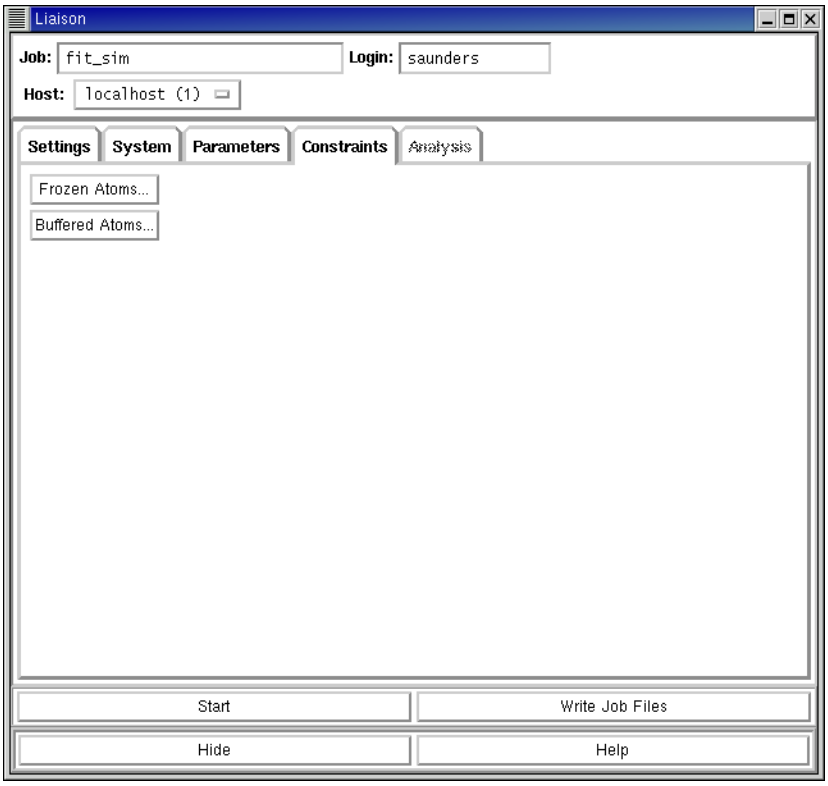


Figure 7.5. The Constraints ***folder of the*** Liaison ***panel.***

Frozen Atoms

Liaison simulations can be performed with some atoms "frozen," so that they never move from their initial position during minimization or dynamics. Clicking this button opens the Frozen Atoms panel, which selects atoms of the receptor to be treated as frozen. The atoms or atom sets can be chosen by picking atoms, residues, or molecules from the Workspace, or by using the Atom Selection dialog box (ASD). See Chapter 2, the *Maestro User Manual*, or the Maestro online help for information on the ASD. For more information about the Frozen Atoms panel, see the Maestro online help as well as Section 9.5.2 on page 161.

Note: These frozen atom selections are keyed to the Workspace structure that you see on screen. Jobs will only include these constraints if they are run on the Workspace structure.

Buffered Atoms

Clicking this button opens the Buffered Atoms panel, which selects atoms to be treated as "buffered." Buffered atoms are allowed to move, subject to harmonic penalty-function restraints that tether them to their initial positions. The Buffered Atoms panel includes atom selection options, like those in the Frozen Atoms panel, and a Buffer Force text box for setting the force constant applied to the buffered atoms. The default buffer force setting is 25.00 kcal/(Å^{2}mol).

A strategy for allowing receptor atoms closest to the ligand to move freely, while restraining (buffering) a range of atoms at intermediate distance from the ligand and fixing (freezing) still more distant atoms, would be to use ASL first to define as buffered all atoms more than 8 Å from the ligand, and then to define as frozen all atoms more than 12 Å from any atom of the ligand. If the ligand was loaded first into the Workspace (and therefore is molecule 1), the Maestro commands needed to carry out this partitioning would be:

```
impactbufferedset beyond 8. mol.n  1
impactfrozenset beyond 12. mol.n  1
```

Because only the last state assigned to an atom is retained, the result will be to buffer atoms between 8 and 12 Å from the nearest atom in the ligand and to freeze atoms greater than 12 Å away.

Note: These buffered atom selections are keyed to the Workspace structure that you see on screen. Jobs will only include these constraints if they are run on the Workspace structure.

7.9 The Analysis Folder

The following analysis jobs are available:

- Determine binding energy model coefficients by fitting to known binding energies (Fit)
- Apply supplied binding model equation coefficients to predict binding energies for new ligands (Predict)

In either case the Simulate task, selected in the Settings folder, must first be used to complete the requisite Liaison runs.

Before an Analysis job can be run, the following conditions must be met:

- Maestro must be in the same directory (Maestro working directory) that was used when the simulation calculations were run.

  You can ensure Maestro is in the correct directory by starting Maestro in the same directory that was used when the simulation calculations were run (the Maestro working directory), or by executing the `changedir` command in Maestro's command input area, to change to the Maestro working directory.

- The Master Liaison Run Directory must be correctly identified.

  You can do this by entering the *jobname* string used in the preceding Liaison simulations in the Job text box at the top of the Liaison panel.

To open the Analysis folder, select Analyze results of earlier simulations in the Settings folder. This selection enables the Analysis folder and disables the System, Parameters, and Constraints folders. The Analysis folder contains two sections: analysis settings and ligand specification.

7.9.1 Analysis Settings Section

The top section of the Analysis folder selects the type of analysis to be performed and, where relevant, takes as input values for the LRM coefficients α, β, and γ, or the GlideScore model coefficients *a* and *b*. The Results of the Analysis display area is also located here.

Analysis type

These options select Fit or Predict as the analysis type:

- Fit. This option requires a text file containing at least three ligand names and the associated binding energies. The specification and format of this file are described below in connection with the Name of Ligand Binding Energy File text box.

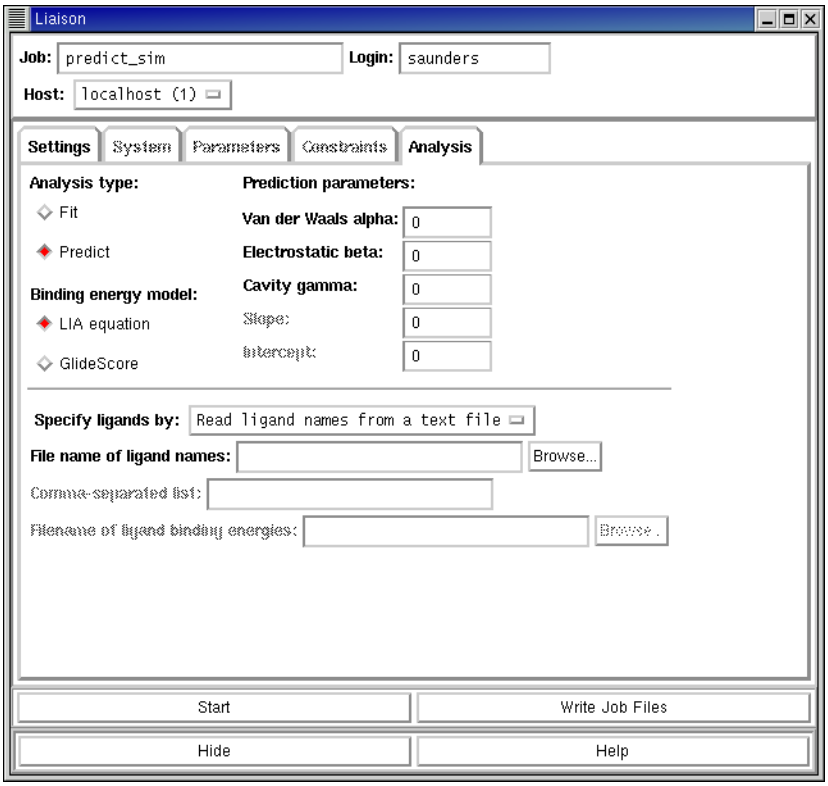


Figure 7.6. The Analysis ***folder of the*** Liaison ***panel.***

- Predict. This option uses values of the selected binding energy model's parameters, together with data retrieved from completed Liaison receptor/ligand simulations, to predict the binding energy for a ligand or a series of ligands. The ligand names can be entered directly or can be taken from an external file. The file format is described under the File Name of Ligand Names feature.

Binding energy model

Use these options to specify which binding energy equation to use for fitting and predictions. The two choices are:

- LIA equation

  This model has three parameters: van der Waals α, electrostatic β, and cavity γ, as described in Section 7.1 on page 113.

- GlideScore

  This is a linear model relating experimental binding energies to the GlideScore. The GlideScore equation given on page 114, therefore, has Slope (a) and Intercept (b) fitting parameters.

Fitting Parameters/Prediction Parameters

These text boxes allow the user to set the coefficients for the selected Binding Energy Model equation that Maestro will use in a Predict calculation. For the LIA Equation, these are the coefficients of the van der Waals, Electrostatic, and Cavity energy terms. For the GlideScore model, they are the Slope and Intercept coefficients. These values are expected to vary for each system studied; there are no "universal" default values.

Parameter Constraints

Fitting jobs will normally derive values for all the parameters appropriate to the selected binding energy model. If you want to set and constrain a value for some, but not, all of the parameters, check the box next to it, then specify a constraint value for the parameter in the Fitting Parameter text box. These boxes do not appear next to parameters not appropriate for the model, and also do not appear for Predict jobs, for which the Analysis folder lists the derived values under Prediction Parameters.

Fit or Predict Output

This read-only text area specifies the name and location of the text file containing the output of the analysis job. The file name will be of the form `liafit_`*jobname*`.out` or `liapredict_`*jobname*`.out`, where *jobname* is the name in the Job text box. This name must be the same as the name of the Master Liaison Run Directory. The output file is written in the Maestro working directory. Multiple Fit or Predict tasks having the same *jobname* will overwrite the previous copy of the output file. To prevent this, move or rename the files from a terminal window.

The Fit or Predict output file can be opened from any available terminal window.

7.9.2 Ligand Specification Section

The bottom section of the Analysis folder identifies the ligands and, for Fit tasks, their binding energies.

Specify Ligands By

This option menu is enabled only for Predict jobs. It allows the ligands to be specified via a user-created text file (the Read Ligand Names from a Text File option) or by entering a comma-separated list of names (the Enter List of Ligand Names option).

File Name

This entry specifies a user-created text file that contains the names of the ligands whose binding energies will be predicted. The names can appear in the file as a comma-separated, space-separated, or carriage-return-separated list. For example:

```
1bkm_3m_1, 1bkm_3m_2, 1bkm_3m_3
```

These entries instruct Maestro to retrieve the requisite data from completed Liaison simulations from directories `1bkm_3m_1/`, `1bkm_3m_2/`, and `1bkm_3m_3/` under the Master Liaison Run Directory, as defined by the *jobname* entered in the Job text box and is positioned under the Maestro working directory. Thus, one such directory path would be:

```
/home/user/liaison/1bkm_3m_1
```

if `/home/user` is the Maestro working directory and `liaison` is the Master Liaison Run Directory. For a small number of ligands, it may be more convenient to use the Comma-separated List option.

Comma-separated List

One or more ligands can be specified by entering a comma-separated list of ligand names into the text box. For example:

```
1bkm_3m_1, 1bkm_3m_2, 1bkm_3m_3
```

Each of these names must correspond to a directory under the Master Liaison Run Directory. See the preceding discussion under File Name. Note that tabs cannot be used to separate the ligand names, though spaces can.

Name of Ligand Binding Energy File

This option is available for the Fit task. It requires a text file containing a list of ligand names and associated binding energies. The ligand binding energies should be in kcal/mol to match the output of the Liaison calculation (other choices, such as kJ/mol or pK_i, will give correct numerical results in the same units, but "kcal/mol" will still be printed). You would create a text file like that shown below. Note that the minimum number of ligands is 3, as there are 3 LRM parameters to fit. In actual practice, binding energies of at least 7 and ideally of 10 or 20 or more ligands should be fit.

```
1bkm_3m_1              -10.5
1bkm_3m_2              -10.9
1bkm_3m_3              -11.9
```

The first column contains the user-defined ligand names, while the second lists the experimental binding affinities. Each name in the first column must correspond to a directory under the Master Liaison Run Directory. See the preceding discussion under File Name. Note that only spaces can be used to separate the ligand names and binding energies.

7.10 Running Liaison as a Stand-Alone Program

You can run Liaison simulations from a terminal window by using (and editing, if desired) files written by Maestro when you select Write Job Files from the lower portion of the Liaison panel. Alternatively, you can set up the directory structure and job files by hand or with an automated script. The assumption is that you have used Maestro to create the necessary directory structure and to write the required job files. If you want to automate the process, use Maestro-written scripts as templates.

To run a Liaison simulation from a terminal window, change to the Maestro working directory and enter the following command at the shell prompt:

`./simulate_`*jobname*

This command runs the `simulate_`*jobname* script that Maestro writes to the Maestro working directory. Several relevant lines from such a script are shown below. The script shows that the job covers 5 ligands (Lig1 … Lig5). Note the use of backquotes:

```
JOB_NAME=liaison
JOB_LIST=`echo Lig1 Lig2 Lig3 Lig4 Lig5`
NPROC=1
```

You could run a specific ligand from the command line by altering the `JOB_LIST` line in this script, or in a copy of it. For example, to run only the `1bkm_3m_2` ligand, the `JOB_LIST` line should appear as:

```
JOB_LIST=`echo 1bkm_3m_2`
```

Provided that the requisite directory, structure, and input files exist, additional ligands can be run by adding to the `JOB_LIST` line.

Liaison simulation jobs, which are more computationally intensive than analysis jobs, can be run on several processors simultaneously by appropriately setting the variable `NPROC` (either in Maestro or by editing the `simulate_`*jobname* file). For `NPROC` > 1, `NPROC` jobs will run concurrently. When one job finishes, another one will start, until all receptor-ligand pair simulations have been submitted. Of course, you need to have sufficient Liaison licenses to run the requested number of jobs.

Liaison Fit or Predict jobs can also be started from the shell prompt by running the `fit_`*jobname* or `predict_`*jobname* script Maestro writes to the Maestro working directory. However, fitting and prediction calculations are virtually instantaneous, once the prerequisite simulation calculations have finished, and it is usually more convenient to submit such jobs directly from the Maestro interface.

7.11 Killing Liaison Jobs

The job control facility may be used to manage and, if necessary, kill Liaison jobs. This facility can be invoked from the Maestro Monitor panel. To kill a Liaison job, select the *jobname*_sim entry and click the Kill button, and all the subjobs will quit as well.

Liaison jobs can also be controlled from the command line with the `jobcontrol` command.

`$SCHRODINGER/jobcontrol` *action* [*job_selection*]

where *action* is one of the following and *job_selection* specifies one or more jobs. The action will be applied to each selected job.

`-list`	List the JobId, job name and status. By default, lists all active jobs.
`-show`	Show basic information about the job
`-kill`	Terminate the job immediately
`-stop`	Terminate the job as soon as possible
`-pause`	Suspend the job temporarily
`-resume`	Continue running a paused job
`-monitor <n>`	Ask for monitoring files to be sent every *n* sec
`-cancel`	Cancel a job that has been launched, but not started
`-purge`	Remove completed job from the database

The *job_selection* argument consists of one or more JobIds, job names, status codes, or queries. This field is optional; if *job_selection* is omitted, the default selection is the query `status!=completed`, that is, all active jobs. It can also be the word `all`, to select all jobs in the jobs database.

For the complete list of job control actions, use the `-help` option to print a usage summary, which also provides a summary of query construction with examples. For more on the job control facility, see the Maestro User Manual.

For the purpose of killing Liaison jobs, the `-list` or `-show` actions can be used to list the jobs in the jobs database, and `-kill` *jobname* or `-kill` *jobid* can then be used to kill one of these jobs. The top-level simulation script is called `simulate_`*jobname*. This script launches a job named *jobname*_sim. Use the JobId corresponding to this job as the argument to `jobcontrol -kill` *jobid*—all the individual ligand sub-jobs are killed as well. The job name *jobname*_sim can be substituted for the JobId in the `kill` command.

Chapter 8: QSite

8.1 Using QSite

QSite performs quantum mechanical/molecular mechanical (QM/MM) calculations, using Jaguar for the QM calculations and Impact for the MM calculations. Ligands and other specified regions of a protein complex can be studied using QM, while MM is used for the rest of the molecule.

At each step of a QM geometry optimization, Impact calculates energy terms for MM-QM region interactions; if MM minimization was also specified, it is also performed at each QM step. The next QM step takes into account the new MM atom distribution and energy terms. If a single-point QM calculation is selected, the current QM/MM energy is calculated without MM minimization.

The speed of QSite is largely determined by the size of the QM region. Therefore there is no advantage to making a smaller model protein. However, please note that the `-s huge` Impact executable option, needed for systems with more than 8000 atoms or 8000 bonds, is not available for QSite jobs.

Cartesian constraints may be placed on atoms in both the QM and the MM regions. See Section 7.8 on page 130 for a brief description of the two types of constraints. Frozen-atom constraints can be applied to atoms in both regions. Buffered-atom constraints can be specified for MM-region atoms, but are ignored if applied to QM-region atoms.

Note: In general, a QSite calculation can only be performed using a single entry. If you want to run a QSite job using the Workspace structure as input, and that structure includes multiple entries, combine them into a single entry using the Merge option from the Entry menu in the Project Table panel. The merged entry should be the only entry included in the Workspace when you start the job. One exception to this is when setting up a transition-state search. In this case you may select up to three Project Table entries, depending upon the algorithm that is selected for performing the search. See Section 8.6 on page 144 for more information about transition-state searching.

The default job name for a QSite job is `qsitetmp`. QSite jobs can be monitored in the job Monitor panel. However, if a QSite job needs to be restarted, it must be restarted from the command line.

To restart a QSite job:

1. Rename the *jobname*`_out.mae` file to *jobname*`.mae`.
2. Rename the *jobname*`.jaguar.01.in` file to *jobname*`.jaguar.in`.

3. Restart the job using the command:

   `impact -i` *jobname*`.inp -j` *jobname*`.jaguar.in`

8.2 The QSite Panel

To open the QSite panel, choose QSite from the Applications menu.

The QSite panel has five tabbed folders, used as follows:

- Potential: choose settings for the MM potential energy function.
- Constraints: set up atom constraints for atoms in the MM and QM regions.
- Minimization: set up energy minimization of the MM region.
- Optimization: select and set up the job task: single-point energy calculation, geometry optimization, or transition-state search.
- QM Settings: specify the QM region and other QM options.

8.3 The Potential Folder

The first three QSite folders, Potential, Constraints, and Minimization, share most of their features with the Potential, Constraints, and Minimization folders in Basic Impact application panels such as Impact Energy Minimization. These shared features are described in Chapter 9, Energy Minimization.

The QSite panel Potential folder contains settings for the MM potential energy function, one of which (continuum solvation) affects the QM potential energy as well.

Note: The following methods and options are not available for QSite calculations: Fast Multipole Method (FMM), periodic boundary conditions and Ewald summation, and treatment of continuum solvation using the Surface Generalized Bohr (SGB) method.

The molecular-mechanics force field used in QSite is the OPLS1999 version of OPLS-AA.

QSite Potential Folder Options

See Section 9.4.1 for a description of the options for Electrostatic treatment and Dielectric constant.

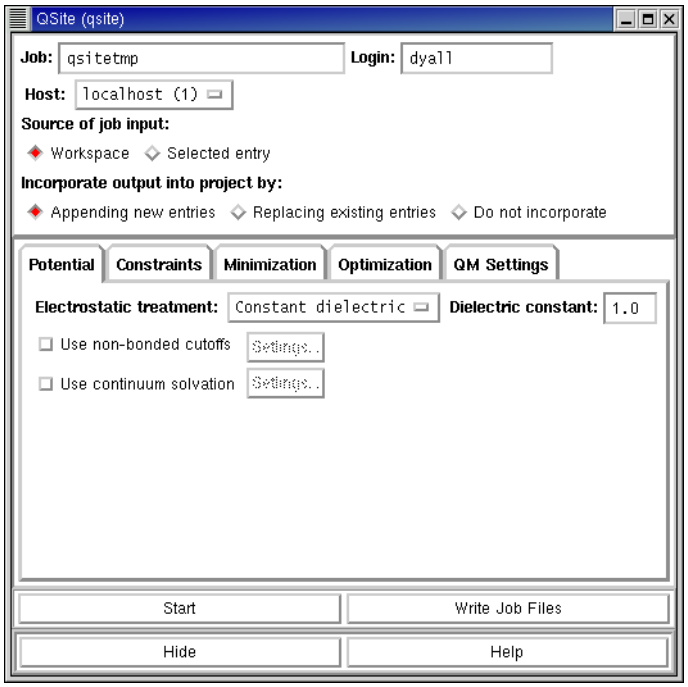


Figure 8.1. The Potential ***folder of the*** QSite ***panel.***

Use distance-based cutoffs

Select this option (previously Use truncation) to truncate nonbonded interactions. (For more information, see Section 9.4.3 on page 157.) There are two settings which can be changed:

- Update neighbor-list frequency: Choose the number of steps after which the neighbor list will be updated. The default is 10 steps.
- Residue-based cutoff distance: All atoms within complete residues which have any pair of atoms within this distance will be included in the nonbonded interaction list. The default is 12 Å.

Use continuum solvation

This option affects both MM and QM calculations. Do not select Use continuum solvation if you would like to run the QM (Jaguar) calculation using multiple processors (parallel processing); when QSite jobs with solvation are run in parallel, erroneous energies result. For more information about continuum solvation options, see Section 9.4.6.

The only available solvation method in QSite, for both the MM and the QM solvation functions, is the Poisson–Boltzmann Solver (PBF); selecting Use continuum solvation

automatically sets both solvation functions to PBF (in the Jaguar input file, `isolv=2.`) The following settings can be changed:

- Resolution: The Poisson–Boltzmann solver involves a finite-element calculation on a grid. The grid spacing controls the accuracy of the PBF calculation and the time required. The default, Low resolution, suffices for most protein work. If needed, greater accuracy can be achieved by choosing Medium or High resolution.
- Displacement threshold: This text box specifies how far (in Å) any atom may move from the coordinates used in the previous PBF calculation before a new PBF calculation must be performed. If no atom has moved this distance, the previously calculated PBF energy and forces are used.

In other respects, the QSite Potential folder offers the same functionality as the Impact Energy Minimization panel Potential folder, which is described in detail in Section 9.4 on page 154.

8.4 The Constraints Folder

The QSite panel Constraints folder is used to apply constraints to the Cartesian coordinates of selected atoms in the MM and/or the QM region. Specified atoms can be frozen at their input coordinates (frozen-atom constraints), or they can be constrained to remain near their initial coordinates by applying a harmonic force (buffered-atom constraints.) In QSite, frozen-atom constraints can be applied to atoms in both the QM and the MM regions. Buffered-atom constraints are used only when they are applied to MM-region atoms; if applied to QM-region atoms, they are ignored.

Atom constraints in QSite, for atoms in the QM as well as the MM region, must be set using the Constraints folder. They *cannot* be set using an `&coord` section in the *jobname*`.jaguar.in` file.

Features of the QSite Constraints folder include two buttons:

- Frozen Atoms
- Buffered Atoms

These buttons open the same panels as the corresponding buttons in the Impact Energy Minimization panel Constraints folder. Both panels are described in detail in Section 9.5 on page 160.

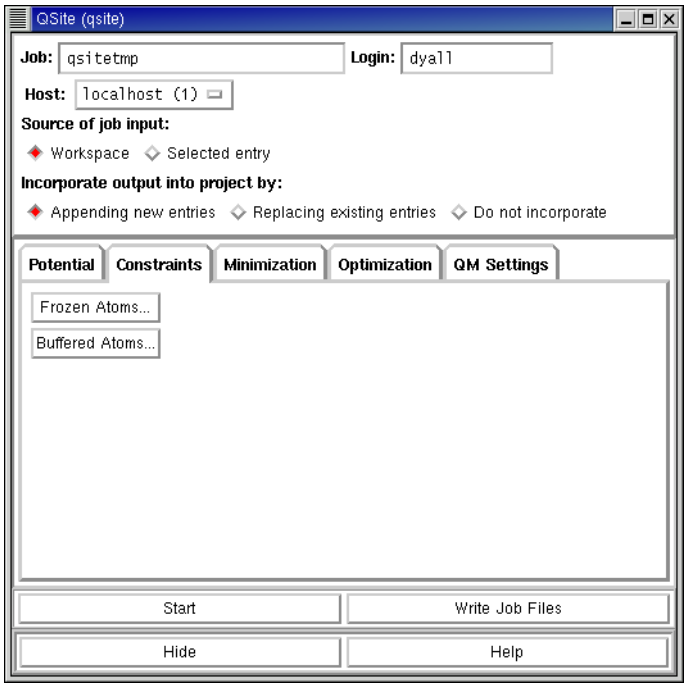


Figure 8.2. The* Constraints *folder of the* QSite *panel.

8.5 The Minimization Folder

The Minimization folder specifies settings for Impact energy minimization of the MM region of the molecule. These settings are not used, and no MM minimization is performed, if the QM method chosen in the Optimization folder is Single point.

The Truncated Newton minimization algorithm is not available for QSite.

In other respects, the QSite Minimization folder offers the same functionality as the Impact Energy Minimization panel Minimization folder, which is described in detail in Section 9.6 on page 162.

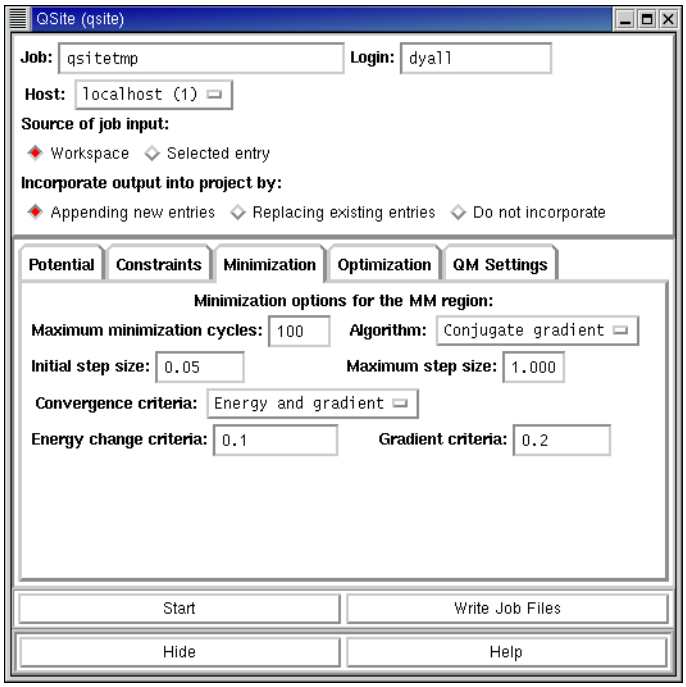


Figure 8.3. The Minimization ***folder of the*** QSite ***panel.***

8.6 The Optimization Folder

The QSite Optimization folder specifies the QM (Jaguar) calculation to be performed and identifies any additional structures that may be needed.

- Single-point energy calculation (no geometry optimization)
- Geometry optimization to a minimum-energy structure
- Geometry optimization to a transition state (TS) using one of three methods
    - Standard
    - Linear Synchronous Transit (LST)
    - Quadratic Synchronous Transit (QST)

Note: QSite QM geometry optimizations use Cartesian coordinates only.

Transition-state optimization by the Linear Synchronous Transit method requires initial guess structures for the reactant and the product. The Quadratic Synchronous Transit method requires initial guesses for the reactant, product, and transition-state structures.

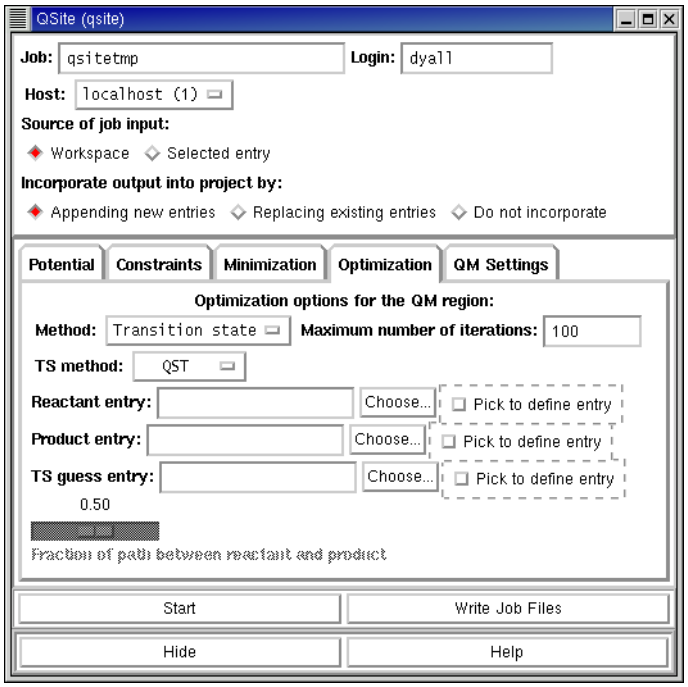


Figure 8.4. The** Optimization **folder with** TS method: QST **selected

These initial guess structures can be selected items in the Workspace or specified entries in the Project Table.

The Optimization folder includes the following menus and options:

Method

The Method menu controls the QM calculation type. The options are:

- Single point (default)
- Minimization
- Transition state

If you choose to run a Single point calculation, the QM energy is calculated for the structure as it stands. No QM geometry optimization *or* MM minimization is performed, and the settings in the Minimization folder are ignored. The other features in the Optimization folder are not needed and are dimmed; go on the QM Settings folder.

Choose Minimization to locate a minimum-energy structure by geometry optimization. If desired, change the default value of 100 for Maximum number of iterations before proceeding to the QM Settings folder. By default both the MM region and the QM region will be optimized. If you want to optimize only the QM region, simply set the number of

minimization steps to 0 in the Minimization folder. There is no need to explicitly freeze all of the MM atoms.

Maximum number of iterations

This box controls the number of optimization iterations for minimization and transition-state calculations. The default is 100 iterations.

TS method

If you select Transition state from the Method menu, the default option for TS method is Standard. The following three methods for transition-state optimization are supported in QSite, corresponding to well-known ab initio techniques. See the *Jaguar User Manual* for detailed information about these methods:

- Standard: The standard transition-state optimization method is useful if you have only a single initial guess structure (the structure in the Workspace) for the transition-state. It attempts to find the saddle point closest to the starting structure by maximizing the energy along the lowest-frequency mode of the Hessian and minimizing the energy along all other modes.
- LST: *Linear Synchronous Transit* is useful if you have initial guess structures for the reactant and the product and want QSite to look for a transition-state structure by interpolating between them. LST uses a quasi-Newton method to search for the optimum transition-state geometry, choosing a transition-state guess structure based on the interpolation value you set using a slider. The default interpolation value is 0.50, directing QSite to choose a transition-state guess structure midway between reactant and product. For a guess structure closer to the reactant, set the interpolation value between 0.0 and 0.50. For a guess structure closer to the product, set the value between 0.50 and 1.0.
- QST: *Quadratic Synchronous Transit* is useful if you have initial guess structures for the reactant, the product, and the transition state. QST uses a quasi-Newton method to optimize the transition-state geometry.

Reactant entry
Product entry
TS guess entry

You can select the initial guess for the reactant structure by typing in the entry name from the current Project Table, by clicking Choose and selecting the entry from a list, or, if the structure is displayed in the Workspace, by selecting Pick to define entry and clicking on any atom in the structure. The same options are available for selecting the Product entry and the TS guess entry.

Fraction of path between reactant and product

When the TS method option menu is set to LST, this slider is available. By default it is 0.50, directing QSite to choose an interpolated transition-state guess structure midway between the reactant and the product. If you want to pick a guess structure closer to the reactant, move this slider between 0.00 and 0.50. For a guess structure closer to the product, select a value between 0.50 and 1.00.

8.7 The QM Settings Folder

The QM Settings folder is used to enter information for the QM job and to define the QM region.

QM job information includes the quantum-mechanical method to be used, the charge and spin multiplicity of the QM system, the number of processors (if Jaguar parallel processing is available), and other keywords and options that may be required by Jaguar.

The QM region can be defined by:

- Selecting the ligand, metal ions, or other disconnected species (not covalently bonded to the protein).
- Specifying *cuts* between certain covalently-bonded atoms in connected residues. QSite cuts are specially parameterized frozen-orbital boundaries between the QM and MM regions.

8.7.1 QM Settings Folder Features

Method

The options for QM method are Density Functional Theory (DFT-B3LYP), Hartree-Fock, and Local Møller-Plessett perturbation theory (Local MP2).

Charge
Multiplicity

The Charge text box should contain the net charge of the QM region of the system, and the Multiplicity text box should contain the associated spin (1 for singlet, 2 for doublet, and so on). Maestro updates the Charge text box with reasonable entries whenever a new residue or ion is added to the QM region. However, both the Charge and Multiplicity text boxes can be edited manually. If the value in the Charge text box does not match the sum of the formal charges of the atoms in the QM region, Maestro displays a warning message, but allows you to proceed. If there is a discrepancy between the total charge and the multiplicity, Jaguar will halt with an error message. The charge and multiplicity of the QM region must be mutually consistent.

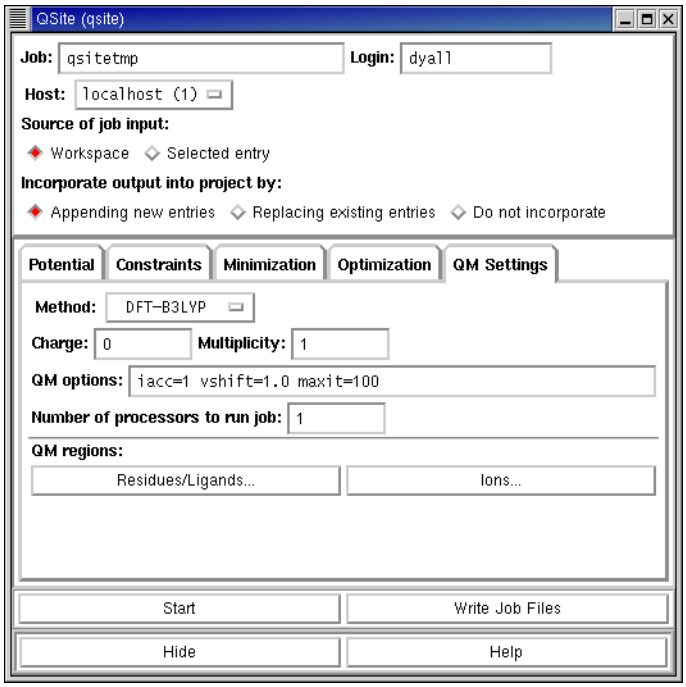


Figure 8.5. The* QM Settings *folder of the* QSite *panel.

QM options

This text box can contain any Jaguar keywords such as print flag settings, non-default convergence criteria, and so on. Each such option is of the form keyword=value (with no embedded blanks). Multiple keyword/value pairs can be specified, separated by one or more blank spaces. By default, the following QM options appear in the box:

```
iacc=1 vshift=1.0 maxit=100
```

You can remove or modify these options as appropriate. See the *Jaguar User Manual* for more information on these keywords.

Number of processors to run job

The QM (Jaguar) portion of the QSite job can be run in parallel if multiple processors are available. Use of this option requires a license for parallel Jaguar. Specify the number of computer processors that will be used for the QM calculation.

Do not use parallel processing for jobs where solvation is selected (Use continuum solvation in the QSite Potential folder.) See Section 8.3.

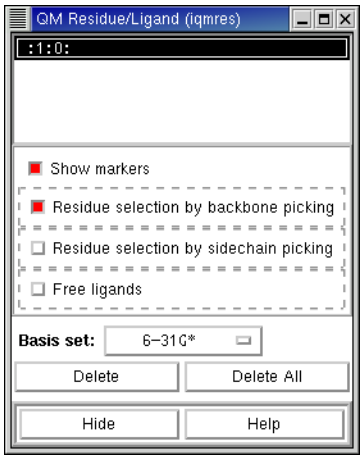


Figure 8.6. The QM Residues/Ligands ***panel in*** QSite**.**

QM regions

These options allow you to define the QM region.

Residues/Ligands

This button opens the QM Residues/Ligands panel, which is used to define and add residues to the QM region (see Section 8.7.2).

Ions

This button opens the QM Ions panel, which is used to add atoms (typically ions) to the QM region (see Section 8.7.3).

8.7.2 The QM Residues/Ligands Panel

The QM Residue/Ligand panel is used to select and add protein residues or non-covalently-bound ligands to the QM region. The features of the QSite Residue/Ligand panel are:

List of QM region residues

The text area at the top of the panel displays the list of QM residues as it is being constructed. Note that Maestro assigns a residue number to every part of the input system. Thus “residue” may refer to either an amino acid residue in the protein, or to a free ligand or ion or solvent molecule.

Show markers

If Show markers is selected, a red trace highlights all of the residues selected for the QM region. Any residue that is part of the QM region will also have its chain name, molecule number, residue number, and insertion code (if applicable) included in the list at the top of the panel.

The QM Residue/Ligand panel supports the following three methods for defining the QM region.

- Residue selection by backbone picking
- Residue selection by sidechain picking
- Free ligands

The first two methods are used to select amino acid residues or side chains. These methods place "cuts" across bonds to alpha carbon atoms in the protein backbone, and these cuts define the QM/MM boundary. The third method, Free ligands (previously called Ligand residue selection) is used for non-covalently-bound structures such as solvent molecules or free ligands. No cuts are made by this method.

To select residues for inclusion in the QM region by backbone picking:

1. Click Residue selection by backbone picking.
2. Cuts must be made at the beginning and end of the QM region. To do this, pick any backbone atom other than an alpha carbon. A cut will be placed between the selected atom and the alpha carbon bonded to it.
3. Pick a second backbone atom. Another cut will be made between this atom and its adjacent alpha carbon, and all of the backbone and side-chain atoms between the two cuts will be included in the QM region. A minimum of three backbone bonds must exist between any pair of backbone cuts. Due to parametrization limitations, backbone cuts cannot be made in glycine (GLY) or proline (PRO) residues, or in residues immediately adjacent to GLY or PRO.

To select a side chain for inclusion in the QM region:

1. Click Residue selection by sidechain picking.
2. Pick any atom from the desired side chain. A cut will be made between the alpha carbon and beta carbon of that residue. All of the atoms in the side chain will be part of the QM region.

Side-chain cuts can be made in any peptide residue other than alanine (ALA), arginine (ARG), glycine (GLY), proline (PRO), serine (SER), and threonine (THR). These residues are excluded because cuts for these side chains are especially difficult to parameterize. The side chains for these residues can still be treated quantum mechanically using backbone cuts to select a QM region that includes the desired side chain.

Cuts in a protein-ligand complex must be between atoms in peptide residues. Covalently-bound ligands can be included in the QM region, but only along with attached protein atoms. The QM region must extend at least as far as the first permissible cut between protein atoms.

To select all atoms in a free molecule for inclusion in the QM region:

1. Click Free ligands.
2. Pick any atom in a free ligand to add all its atoms to the QM region.

This selection is used to add an entire ligand, or other non-covalently-bound species, to the QM region. Free atoms, such as metal ions, are selected using the Ions panel. Ligands which are covalently bound to the protein cannot be added using this method, because this method does not make any parametrized cuts. To add covalently bound ligands to the QM region, make either a pair of backbone cuts to select the residue to which the ligand is bound, or make a side-chain cut.

Basis set

By default, the basis set used for the entire QM region is 6-31G* (LACVP* in the case of transition metals), which is the basis set used in developing the parametrizations for the cuts. If you wish to use a different basis set for any residue in the QM region, select that residue in the QM region list, then select another basis set from the Basis set menu.

Delete

To remove a specific residue from the QM region, highlight the residue in the QM region list and click Delete at the bottom of the panel. When deleting individual residues, take care to ensure that the resulting QM region is consistent with the QSite cuts previously made using the Residue selection by backbone picking or Residue selection by sidechain picking buttons.

Delete All

Click Delete All to remove all residues from the QM region.

8.7.3 The QM Region Ions Panel

The QM Region Ions panel is used to include free atoms (typically ions) in the QM region of a QSite job.

Note: This option should not be used on covalently-bonded atoms, such as individual protein atoms, and should not be used to select a subset of ligand atoms, because valid QM/MM cuts would not be obtained.

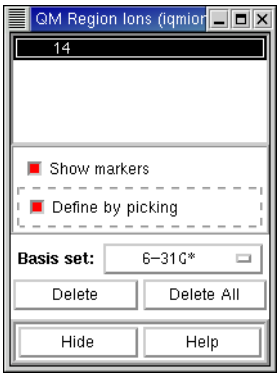


Figure 8.7. The QM Region Ions ***panel in*** QSite**.**

Click Ions in the Settings folder to open the QM Region Ions panel. The selection options in the QM Region Ions panel are:

QM Region Ions list

The text area at the top of the panel displays the list of QM-region ions as it is being constructed.

Show markers

If this option is selected (the default), a set of green crosses will highlight the ions picked for the QM region.

Define by picking

Selected by default. Clicking on a free atom or ion in the Workspace adds that atom to the QM region.

Basis set

Once an ion has been added to the QM region, the basis set to be used in the calculation can be changed on a per-atom basis. To do so, click on the ion in the ions list, then select another basis set from the Basis set menu.

Delete

To remove a specific ion from the QM region, select the ion in the ions list and click Delete.

Delete All

Click Delete All to remove all ions from the QM region.

Chapter 9: Energy Minimization

9.1 Basic Impact Applications

Basic Impact Applications are general-purpose molecular mechanics simulations that you can launch from Maestro or from the command line, as described in Chapter 3. There are four Basic Impact applications:

- Energy Minimization
- Molecular Dynamics
- Hybrid Monte Carlo
- Soak

which will be described in the following chapters.

For an extensive set of examples of input files for Basic Impact applications, see Appendix C of the *FirstDiscovery Command Reference Manual*. The *Command Reference Manual* is provided on disk with the FirstDiscovery distribution and is also available at Schrödinger's support webpage, http://www.schrodinger.com/Support/pdf.html

9.2 Using the Energy Minimization Panel

The Impact Energy Minimization panel is used to set up and run an Impact energy minimization calculation on the structure in the Workspace.

To open the Impact Energy Minimization panel:

- In the Maestro Applications menu, choose Minimization from the Impact submenu.

In the upper part of the panel are the standard FirstDiscovery panel options for Job name, Login, and Host, as well as Source of job input and Incorporate output into project by. For a description of these options, see Section 1.2. The default job name for Impact Energy Minimization jobs is `impacttmp`.

The input structure for an Impact Energy Minimization job can be either the contents of the Workspace or a single entry in the Project Table.

To perform energy minimization on a system composed of multiple entries:

1. Include only those entries in the Workspace.
2. Select Workspace as the Source of job input.

9.3 Energy Minimization Panel Features

The Energy Minimization panel has three tabbed folders:

- Potential
- Constraints
- Minimization

All three folders are described in detail in this chapter. The QSite panel includes versions of these three folders. The first two folders also appear in the Maestro panels of two other Basic Impact applications: Molecular Dynamics Simulations and Hybrid MC Simulations.

9.4 The Potential Folder

The Potential folder sets parameters that control how Impact calculates the molecular-mechanics energy in a minimization calculation or dynamics simulation. The panels for three Basic Impact Applications and for QSite each include a Potential folder. Open the the folder by clicking the Potential tab, which is always the first tab from the left.

9.4.1 Potential Folder Options

The options in the upper portion of the Potential folder are described below.

Force field

This option menu sets the molecular-mechanics force field used by Impact. The OPLS-AA all-atom force field is the only choice for FirstDiscovery and Impact calculations. For Liaison and Basic Impact calculations set up from Maestro, only the OPLS-AA 2001 force field is supported. For QSite calculations set up from Maestro, only the OPLS-AA 1999 force field is supported. See the *FirstDiscovery Command Reference Manual* for information about options from the command line.

Parameter file

This text box allows you to specify a file containing the molecular-mechanics parameters to be used by Impact. However, you will usually not need to change the default value. The default molecular-mechanics parameters are stored in a file called `paramstd.dat` that is provided in the FirstDiscovery distribution.

Electrostatic treatment

This option menu offers two methods for calculating the electrostatic component of the molecular mechanics energy:

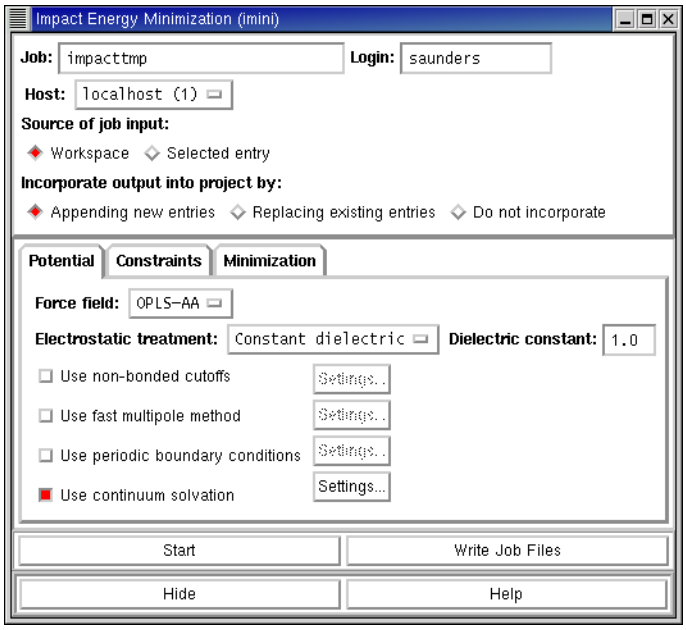


Figure 9.1. The Potential ***folder of the*** Impact Energy Minimization ***panel.***

- Constant dielectric

  This option calculates the electrostatic interaction between atoms *i* and *j* as:

  $E_{ele} = 332.063762\ q_i q_j/(\varepsilon\ r_{ij})$

  where:

  E_{ele} is the electrostatic interaction in kcal/mol

  q_i and q_j are the partial atomic charges on atom *i* and *j*

  r_{ij} is the distance in Å between atoms *i* and *j*

  ε is the Dielectric constant (see below)

  A constant dielectric is appropriate for a vacuum (gas-phase) calculation or when an explicit or implicit solvent model is used.

- Distance dependent dielectric

  This option calculates the electrostatic interaction between atoms *i* and *j* as:

  $E_{ele} = 332.063762\ q_i q_j/(\varepsilon\ r_{ij}^2)$

  where:

  - E_{ele} is the electrostatic interaction in kcal/mol
  - q_i and q_j are the partial atomic charges on atom *i* and *j*

- r_{ij} is the distance in Å between atoms *i* and *j*
- ε is the Dielectric constant (see below)

A distance-dependent dielectric is sometimes used as a primitive model for the effect of solvent. In this model, the electrostatic interaction between a pair of atoms falls off rapidly as the distance between the atoms increases. However, continuum and explicit solvent models are much better at accounting for solvent effects than a distance-dependent dielectric.

Dielectric constant

This text box specifies the value of the dielectric constant ε used in the electrostatic calculations.

9.4.2 Potential Folder Methods

The lower part of the Potential folder allows you to choose among molecular mechanics treatments. When the check box for a method is selected, clicking the associated Settings button opens a panel of relevant options. The methods are listed briefly in this section, and then each Settings panel is described in more detail.

Use truncation

In molecular-mechanics calculations it is often impractical to include the nonbonded (electrostatic and van der Waals) interactions between every pair of atoms. For large systems, many such pairs are separated by a great distance and contribute little to the overall interaction energy. Judicious use of truncation to remove interactions between widely separated pairs of atoms is an important strategy for reducing the time and memory required to perform calculations on large systems.

Use fast multipole method

The Fast Multipole Method (FMM) is an algorithm for speeding up the electrostatic part of the molecular-mechanics calculation for large systems.

Use periodic boundary conditions

Periodic boundary conditions are commonly used for calculations with explicit solvent, but can be employed for any periodic system.

Use continuum solvation

Two implicit solvent models, the Surface Generalized Born Model (SGB) and the Poisson-Boltzmann Solver (PBF), are available in Impact. These methods account for the effects of solvent without the use of explicit water molecules.

9.4.3 The Truncation Panel

The Truncation panel defines the truncation settings for an Impact calculation. When sufficient care is taken, the use of truncation to remove interactions between widely separated pairs of atoms is an important strategy for reducing the time and memory required to perform calculations on large systems. Currently only residue-based cutoffs are supported for calculations set up by Maestro. This means that all atoms within complete residues that have any pair of atoms within the cutoff distance will be included in the nonbonded interaction list.

To open the Truncation panel, select the Use truncation option and click the adjacent Settings button. The selection options in the Truncation panel are:

Update neighbor-list frequency

When truncation is active, all the pairs that fall within the cutoff radius are stored in a "neighbor list." During a minimization calculation or a dynamics simulation, the geometry of the structure may change so as to bring some pairs of atoms that were originally outside the cutoff distance to within the cutoff. Conversely, some pairs of atoms may move outside the cutoff distance. For these reasons, the neighbor list needs to be updated from time to time. The frequency of this update is controlled by this integer field. By default the neighbor-list is updated every 10 minimization or dynamics steps. Increasing this value (updating the neighbor-list *less* often) will speed the calculation but may affect the accuracy of the results. Decreasing this value (updating the neighbor-list *more* often) will slow the calculation but may improve the accuracy.

Residue-based cutoff distance

This text box specifies the value for the cutoff distance. Increasing the cutoff distance will slow the calculation and require more memory, but may yield more accurate results. Decreasing the cutoff will speed up the calculation, but may reduce the accuracy of the results if significant nonbonded interactions are omitted. This is especially true for systems that include formally charged atoms, as such systems can have large long-range electrostatic interactions.

9.4.4 The Fast Multipole Method Panel

For large systems, the Fast Multipole Method (FMM) speeds the evaluation of the electrostatic and van der Waals parts of the molecular-mechanics energy by using interacting hierarchical multipoles to approximate the true electrostatic potential.

When used with periodic boundary conditions (the box must be cubic), the FMM method requires that the net charge of the system be zero. FMM calculations also can be carried out on isolated, non-periodic systems, but this is seldom cost-effective except for systems containing tens of thousands of atoms.

Note: When Use fast multipole method is selected, the Truncated Newton (TN) minimization method and the SGB continuum solvation method are unavailable.

To open the Fast Multipole Method panel, select Use fast multipole method and click the adjacent Settings button. The following options can be set:

Level

Use this text box to set the number of levels in the hierarchical tree used in the FMM calculation. This setting is relevant only when the Reversible RESPA integration propagator is used with more than two stages. (See Section 10.4 on page 167.)

The Level parameter specifies the number of times the elementary simulation box is divided into halves along each direction, a procedure known as octree decomposition. Thus, if Level = 1 is set, one division is made along X, one along Y, and one along Z, so that the box is divided into eight sub-cubes (octants). If Level = 2, each sub-cube is further divided into eight smaller cubes, for a total of 64, and so on. Set Level to at least 2. Larger values result in increased accuracy at the cost of longer execution time, but they may be useful in very large systems.

Maximum

Use this text box to set the maximum number of multipole moments to be used to approximate the potential and field produced by "far" clusters of atoms. Currently a minimum of 4 and a maximum of 20 multipoles are permitted.

Use smoothing

If you select this option, a smooth cutoff is used to separate into "near" and "far" components the forces that are computed explicitly from Coulomb's Law rather than from the multipole expansions. This setting is relevant only when the Reversible RESPA integrator is used with more than two stages. (See Section 10.4 on page 167.)

9.4.5 The Periodic Boundary Conditions Panel

Impact calculations can be performed with periodic boundary conditions. This technique is usually applied with explicit solvent in order to avoid nonphysical "edge effects." The system of interest is defined to be in a box of a given size, images of which are replicated throughout space to form an infinite 3D lattice.

To open the Periodic Boundary Conditions panel, select Use periodic boundary conditions, then click Settings. The selection options of the Periodic Boundary Conditions panel are:

Box X, Y and Z lengths

Use these three text boxes to set the size of the simulation box. The minimum size for any dimension that Maestro will use is 18.62 Å.

Use Ewald long-range correction

Click to select the Ewald summation method for efficiently summing long-distance electrostatic interactions in periodic systems. Unlike the Fast Multipole Method, Ewald does not require the net charge of the system to be zero. This setting is ignored when the Fast Multipole Method is used.

- Maximum length of K-space vectors

  The value in this text box is used to determine the number of component terms retained in the reciprocal-space part of the Ewald summation. The default value is 5. Larger values yield increased accuracy but result in slower execution.

- Alpha

  This text box sets the value of the parameter alpha in the Ewald method. A reasonable value is 5.5/L, where L is the linear dimension of the cubic simulation box. The default value is 0.25.

9.4.6 The Continuum Solvation Panel

Impact supports two implicit solvent models, the Surface Generalized Born Model (SGB) and the Poisson-Boltzmann Solver (PBF). These methods simulate the effects of solvent without the use of explicit solvent molecules.

To open the Continuum Solvation panel, select Use continuum solvation and click Settings. The selection options of the Continuum Solvation panel are:

Solvation Method

The Solvation Method menu options are:

- Surface Generalized Born Model (SGB), the default, which is unavailable if Use fast multipole method has been selected.
- Poisson Boltzmann Solver (PBF), which is unavailable if the Truncated Newton minimization algorithm has been selected.

SGB Displacement Threshold

This text box specifies how far (Å) any atom may move from the coordinates used in the previous SGB calculation before a new SGB calculation must performed. If no atom has moved this distance, the previously calculated SGB energy and forces are used.

PBF Resolution

The Poisson-Boltzmann solver involves a finite-element calculation on a grid. The grid spacing controls both the accuracy of and time required for the PBF calculation. The

default is to use a Low resolution grid, which should suffice for most protein work. If needed, greater accuracy can be achieved by setting this option menu to Medium or High.

PBF Displacement Threshold

Use this text box to specify how far (Å) any atom may move from the coordinates used in the previous PBF calculation before a new PBF calculation must be performed. If no atom has moved this distance, the previously calculated PBF energy and forces are used.

9.5 The Constraints Folder

The Constraints folder is used to set up Impact *atom constraints* and *bond constraints*.

- **Bond constraints**: Molecular dynamics and hybrid Monte Carlo simulations can use the SHAKE/RATTLE algorithm to constrain bond lengths. When setting up these jobs from the GUI, you can choose to constrain all bonds (the default for MD) or not to constrain any bonds (the default for HMC.) These options are discussed in Section 10.3 on page 166.
- **Atom constraints**: Impact Energy Minimization, Molecular Dynamics, and Hybrid MC jobs, as well as Liaison simulations and QSite calculations, can include atom constraints. Specified atoms can be frozen at their input coordinates (frozen-atom constraints), or harmonic force constraints (buffered-atom constraints) can be applied to keep them near their input positions.

The QSite panel Constraints folder can be used to specify frozen-atom constraints on atoms in both the QM and the MM regions, but buffered-atom constraints are applied only in the MM region.

The Liaison panel Constraints folder allows atom constraints to be applied to Liaison simulations. See Section 7.8 on page 130.

Note: The Constraints folder in the Glide panel is used to set up *Glide constraints* (required interactions between ligand and receptor atoms) *only*. Impact atom constraints are not used in Glide.

9.5.1 Constraints Folder Features

The options that are not available for Impact Energy Minimization (and therefore not in the Constraints folder when it is opened from the Energy Minimization panel) are described in Chapter 10.

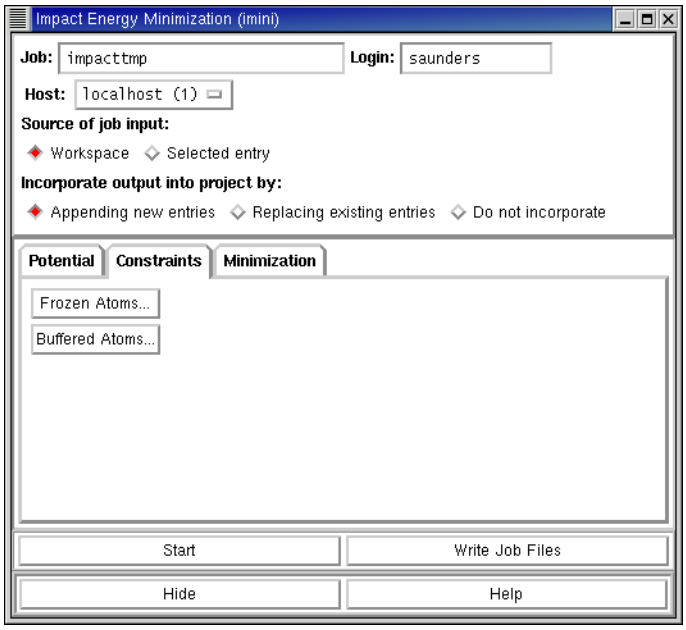


Figure 9.2. The Constraints ***folder of the*** Impact Energy Minimization ***panel.***

Frozen Atoms

Impact calculations can be performed with some atoms completely "frozen," so that they never move from their initial positions during minimization or dynamics. Clicking the Frozen Atoms button opens the Frozen Atoms panel, which selects the atoms to be treated as frozen. The Frozen Atoms panel is described on page 161.

Buffered Atoms

This button opens the Buffered Atoms panel. Use this panel to select the atoms to be treated as "buffered." Buffered atoms are allowed to move, subject to harmonic penalty-function restraints that tether them to their initial positions. The Buffered Atoms panel is described on page 162.

9.5.2 The Frozen Atoms Panel

Use the Frozen Atoms panel to specify a set of atoms to be frozen during an Impact minimization calculation or dynamics simulation. Open the Frozen Atoms panel by clicking the Frozen Atoms button in the Impact Constraints folder.

To define frozen atoms, use the Pick options, the All button, or the Atom Selection dialog box (click the blue Select button). Selected atoms are listed at the top of the Frozen Atoms

panel. If Show Markers is selected, the atoms to be frozen are marked with a red padlock in the Workspace. These markers can be hidden by deselecting Show Markers in the ASD. The Delete button at the bottom of the panel removes the currently selected frozen atom from the list. The Delete All button removes all currently defined frozen atoms from the list.

Note: These frozen atom selections are keyed to the Workspace structure that you see on screen. If you choose Selected entry instead of Workspace as the Source of job input, your frozen atom set is not used.

9.5.3 The Buffered Atoms Panel

Use the Buffered Atoms panel to specify a set of atoms to be harmonically restrained during an Impact minimization calculation or dynamics simulation. Such atoms are referred to as "buffered" atoms because they are often used as a "buffer" between totally free and totally frozen regions. The Buffered Atoms panel is opened by clicking the Buffered Atoms button in the Constraints folder.

The upper portion of the Buffered Atoms panel displays the list of atom numbers that are to be buffered (restrained with a harmonic potential).

Below this list is the Buffer Force text box, which sets the force constant to be applied to the buffered atoms. The default buffer force setting is 25.00 kcal/(Å^{2}mol).

To define buffered atoms, use the Pick options, the All button, or the Atom Selection dialog box (click the blue Select button). Selected atoms are listed at the top of the Buffered Atoms panel. If Show Markers is selected, the atoms to be buffered are marked with a blue cross and a "spring" icon in the Workspace. These markers can be hidden by deselecting Show Markers in the ASD. The Delete button at the bottom of the panel removes the currently selected buffered atom from the list. The Delete All button removes all currently defined buffered atoms from the list.

Note: These buffered atom selections are keyed to the Workspace structure that you see on screen. If you choose Selected entry instead of Workspace as the Source of job input, your buffered atom set is not used.

9.6 The Minimization Folder

The basic settings of the Impact energy minimization task are defined in the Minimization folder.

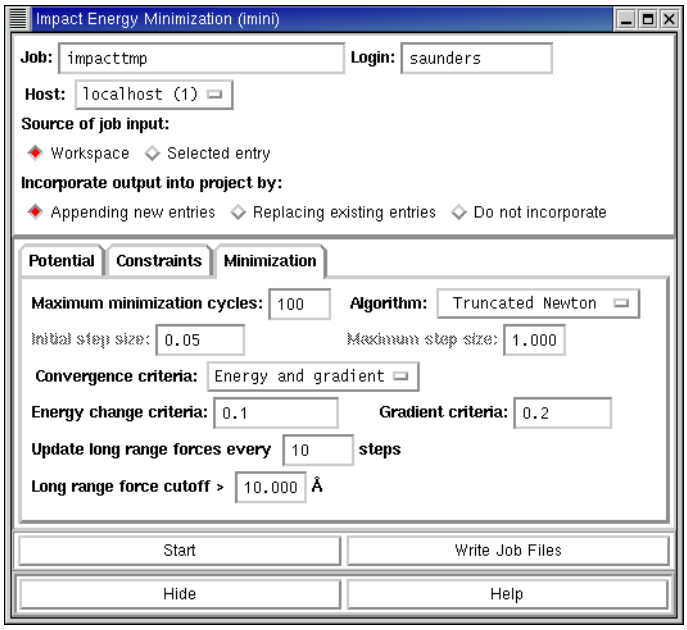


Figure 9.3. The Minimization ***folder of the*** Impact Energy Minimization ***panel.***

Maximum minimization cycles

This text box sets the maximum number of cycles for the minimization calculation. The minimization terminates if it has not converged by this point. The default value is 100 iterations, but you can specify any value greater than or equal to zero. “Zero cycles” is a special case: it instructs Impact just to evaluate the energy for the current coordinates.

Algorithm

This option menu selects the minimization algorithm. The choices are:

- Truncated Newton (TN). This is a very efficient method for producing optimized structures and is also the default. A short conjugate gradient pre-minimization stage is performed first to help improve the convergence of the Truncated Newton algorithm.
    - The Truncated Newton minimization algorithm is not available if the Fast Multipole Method (FMM) has been selected in the Potential folder.
    - The Truncated Newton minimization algorithm is not available if the Poisson-Boltzmann solver (PBF) continuum solvation method has been selected.
- Conjugate gradient. This is a good general optimization method.

- Steepest descent. This can be a good method for initiating a minimization on a starting geometry that contains large steric clashes. Convergence is very poor towards the end of minimization, where the conjugate gradient algorithm should be used.

Initial step size

This text box specifies the initial step size of the minimization cycle. The default value is 0.05 Å, but any positive value is allowed.

Maximum step size

This text box specifies the maximum step size of the minimization cycle. If the step size exceeds this value, the minimization will halt. The default value is 1.00 Å, but any positive value is allowed. The maximum step size is the maximum displacement allowed for an atom in any step of a minimization calculation.

Convergence criteria

This option menu sets the convergence criteria for the minimization. Either or both of two criteria—energy change and gradient—can be specified. Thus, the options are:

- Energy and Gradient. Choosing this option allows access to both the Energy change criteria and Gradient criteria text boxes.
- Energy change criteria. Use this text box to specify the value of the energy change criterion. The default value is 10^{-7} kcal/mol, but any positive value is allowed. The criterion is satisfied if two successive energies differ by less than the specified value.
- Gradient criteria. Use this text box to specify the value of the gradient criterion. The default value is 0.01 kcal/(mol∗Å), but any positive value is allowed. The criterion is satisfied if the norms of two successive gradients differ by less than the specified value.

Long Range Forces Options (for Truncated Newton minimizations):

- Update long range forces every *n* steps. This option is available only for the Truncated Newton algorithm. The default is to update long range forces every 10 steps. Between updates, estimates of these forces are used. Smaller values of *n* (more frequent updates) can be used to improve convergence, but will make the optimization slower. The maximum recommended value is 20.
- Long range force cutoff > *n* Angstroms. This option is available only for the Truncated Newton algorithm, and specifies the distance beyond which forces will be treated as “long range,”—that is, updated every *n* steps as specified in the previous option, and estimated between updates.

Chapter 10: Molecular Dynamics Simulations

10.1 Using the Dynamics Panel

Use the Impact Dynamics panel to set up and run an Impact Molecular Dynamics (MD) simulation on the Workspace structure. Molecular Dynamics simulations examine stable, ground state molecules by applying Newton's equations of motion. The constant volume and temperature (NVT) ensemble is the default ensemble for MD simulations. The constant volume and energy (NVE) and constant pressure and temperature (NPT) ensembles are also supported.

NPT simulations require the use of periodic boundary conditions (see Section 9.4.5 on page 158). Such calculations often, but not always, use explicit solvent (see Chapter 12).

To open the Impact Dynamics panel:

- In the Maestro Applications menu, choose Dynamics from the Impact submenu.

In the upper part of the panel are the standard FirstDiscovery panel options for Job name, Login, and Host, as well as Source of job input and Incorporate output into project by. For a description of these options, see Section 1.2. The default job name for Impact Molecular Dynamics jobs is `impacttmp`.

The input for an Impact Molecular Dynamics job can be either the contents of the Workspace or a single entry in the Project Table.

To perform Molecular Dynamics simulations on a system composed of multiple entries:

1. Include only those entries in the Workspace.
2. Select Workspace as the Source of job input.

10.2 Dynamics Panel Features

The Dynamics panel has four tabbed folders:

- Potential
- Constraints
- MD Parameters
- Dynamics

The Potential and Constraints folders are described in Chapter 9. The Constraints folder in the Dynamics panel includes features not available for Energy Minimization, Liaison, or QSite. These are described here, along with the MD Parameters folder and the Dynamics folder.

10.3 The Constraints Folder

In addition to the features described in Section 9.5 on page 160, the following features appear in the Constraints folder for Impact Dynamics and/or HMC calculations:

Constrain all bonds

This option is selected by default for Impact Dynamics, but is not the default for HMC simulations. When selected, it constrains all bond lengths to the "ideal" values defined by the molecular force field. For HMC and MD simulations, SHAKE is used to constrain the bond length and RATTLE is used to remove the relative motion (velocity) of the bonded atoms along the interatomic axis.

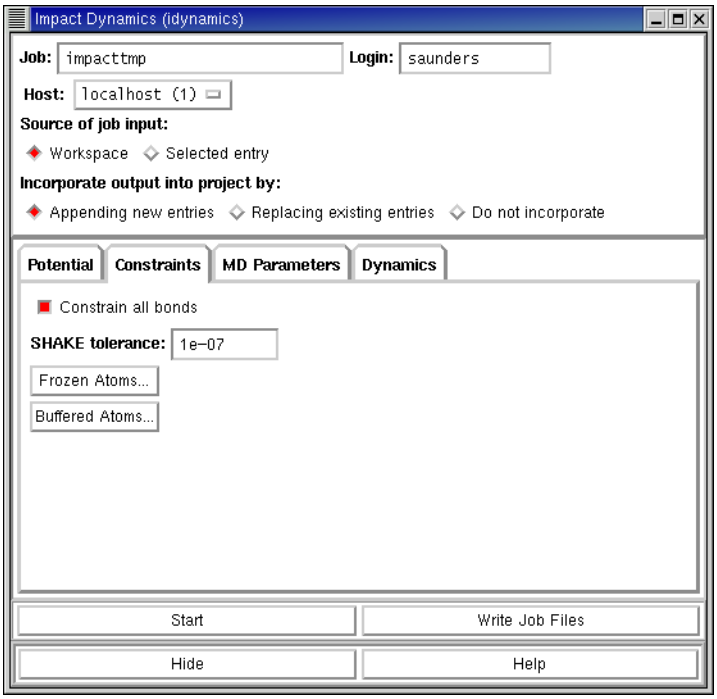


Figure 10.1. The Constraints ***folder of the*** Impact Dynamics ***panel.***

SHAKE tolerance

This text box sets the tolerance for the SHAKE/RATTLE algorithm. The default value is 10^{-7} Å for SHAKE and 10^{-7} Å/ps for RATTLE. Increasing the tolerance will speed the calculation at the cost of allowing greater variation from the ideal values. Any value greater than zero is allowed. Note that the same numeric value is used for both tolerances, even though the units are different.

10.4 The MD Parameters Folder

Use the MD Parameters folder to specify molecular dynamics settings that affect both Molecular Dynamics (MD) and Hybrid Monte Carlo (HMC) calculations (see Chapter 11). Some selection options appear dimmed when they are not relevant for other, previously chosen options.

The selection options are:

Integration algorithm

This option menu specifies the integration algorithm employed to integrate the Equations of Motion (EOM). The options are Verlet and RRESPA. There are no other settings to specify for the Verlet option. For RRESPA, three other text boxes become active, as noted below.

- Verlet. The widely used velocity Verlet integration algorithm is the default for integrating the equations of motion in standard Cartesian-space molecular dynamics.
- RRESPA. The Reversible REference System Propagator Algorithm), the other choice offered by Impact for integrating the EOM, can be much more efficient. By breaking up the integration into large, medium, and small time steps (see the RRESPA update frequencies text boxes), this integrator devotes appropriate computational power to specific classes of forces—and thus to keep the calculation from being dominated by the small time steps needed to accurately integrate the fast motions (such as bond stretches). In particular, RRESPA integrates the fast motions with small time steps and the slow motions (far more numerous) with larger time steps.
- RRESPA update frequencies. If you select RRESPA, text boxes for Fast forces, Medium forces, and Slow forces are enabled.

  When the Fast Multipole Method (FMM) is also used, the forces are separated into three groups: those arising from well-separated bodies, those arising from first and second neighbors that are not very close, and those coming from the local expansions, which include bonded terms.

Table 10.1. RRESPA Text Boxes.

Force	Interacting Species	Default Setting
Fast	Bonded and short-distance electrostatics	4
Medium	1^{st} and 2^{nd} neighbors that are not close	2
Slow	Long-distance nonbonded	1

All text boxes have acceptable ranges of any integer value greater than one. These entries modify the time step for the underlying MD or HMC simulation in the following way. Suppose that the global Time step specified in the Dynamics folder of the Impact Dynamics panel or the Hybrid MC folder of the Impact Hybrid Monte Carlo panel is δ. Then the time step used to integrate the slow forces is $\delta/1$, while the time step for medium forces is $\delta/(1 \times 2)$, and that for fast forces is $\delta/(1 \times 2 \times 4)$. Thus, the integration time step decreases as the product of the cumulative RRESPA update frequencies in going from slow to fast forces.

When FMM is not used (this is the more common case), only the Fast forces text box affects the calculation. In this case, the Medium and Slow forces are combined and use the global Time step set in the Dynamics folder of the Impact Dynamics

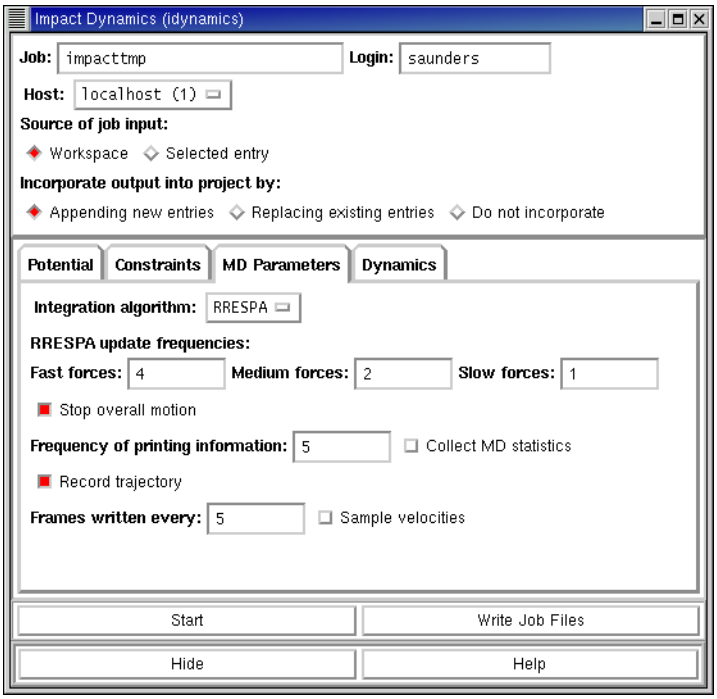


Figure 10.2. The MD Parameters ***folder of the*** Impact Dynamics ***panel.***

panel or in the Hybrid MC folder of the Impact Hybrid Monte Carlo panel. Fast forces use the shorter time step computed by dividing the Time step by the integer entry in the text box.

Stop overall motion

When this option is selected (the default), overall rotational and translational motion (drift) of the system is subtracted from the calculation.

Frequency of printing information

This text box selects the frequency with which MD information is written during the simulation. The default value is to print information every 5 MD steps. Any integer value greater than zero is allowed.

Collect MD statistics

This option is off by default. When it is selected, the MD statistics are collected and are written to the end of the Impact output file. These statistics measure fluctuations of the different energy terms.

Record trajectory

This option is off by default. When it is selected, trajectory information is written to the file *jobname*`.trj` in the Maestro working directory (or for Liaison, in the individual ligand directories arrayed under the Master Liaison Run Directory). This information is written in binary format, but can be analyzed using the `ANALYSIS` task of Impact. A trajectory file contains a sequence of snapshots of the coordinates of the system and, if requested, of the velocities as well.

Note: The Impact analysis task must be run using Impact from the command line. See Chapter 3 for a brief overview of command-line Impact. See the *FirstDiscovery Command Reference Manual* (available on the FirstDiscovery installation disks or on Schrödinger's support webpage, http://www.schrodinger.com/Support/pdf.htm) for additional information.

The following two options are applicable when Record Trajectory has been selected:

Frames written every

This text box specifies how often trajectory information is written to the trajectory file. The default is every 5 MD steps. Any integer value greater than zero is allowed.

Sample velocities

This option is off by default. When it is selected, velocity information is written to the trajectory file.

10.5 The Dynamics Folder

Use the Dynamics folder to set up the type of MD simulation to be performed. Options appear dimmed when they are not applicable for other options already chosen. To open the Dynamics folder, click the Dynamics tab of the Impact Dynamics panel.

The major selection options are described below. The other options on this panel are described under the major option to which they apply.

Number of MD steps

This text box sets the number of MD steps to be used for the simulation. The default setting is 100 steps, but any number greater than zero is allowed.

Time step

This text box sets the time step for the MD simulation. The default value is 0.001 ps, but any value greater than zero is allowed. A somewhat larger value (0.0015 or 0.002) may be suitable if bond lengths are constrained (see the Impact Constraints folder in Section 9.5 on page 160) or if the RRESPA integrator is employed (see the MD Parameters folder in Section 10.4 on page 167).

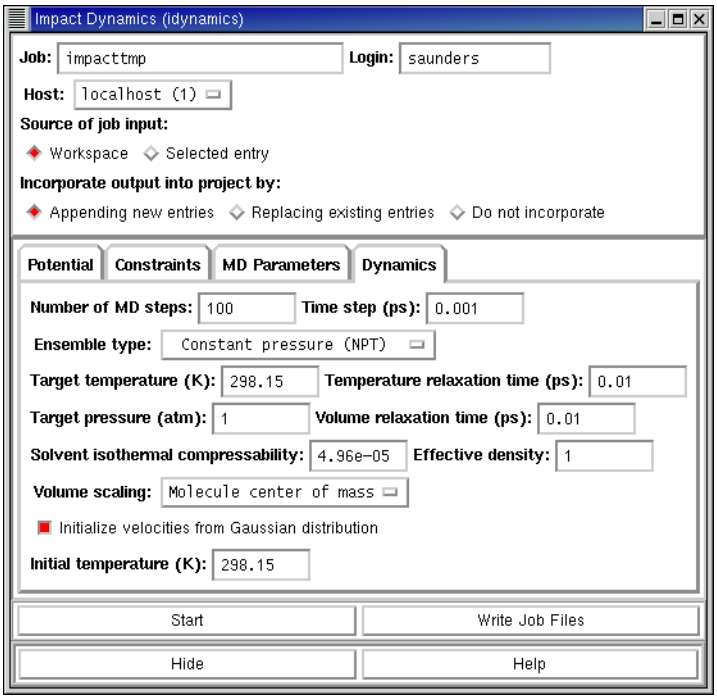


Figure 10.3. The Dynamics ***folder of the*** Impact Dynamics ***panel.***

Ensemble type

Impact offers three choices:

- Constant temperature (NVT)
- Constant energy (NVE)
- Constant pressure (NPT)

Depending on the ensemble chosen, various subsidiary settings become active.

Constant temperature (NVT)

With this ensemble type, volume and temperature are held constant during the simulation. This selection results in coupling the system to an external heat bath with a target temperature that is the same for all molecular species. Two settings become available when this ensemble type is chosen:

- Target temperature. This text box sets the target temperature for a NVT (or NPT) simulation. The actual temperature will fluctuate about the target value. At each MD step the velocities will be scaled so that the temperature will approach the desired value on a timescale determined by the Temperature relaxation time parameter. The default temperature is 298.15 K. The acceptable range is any value greater than or equal to 0 K.
- Temperature relaxation time. This text box sets the temperature relaxation time in picoseconds for velocity scaling. The default value is 0.01 ps. The acceptable range is any value greater than 0 ps.

Constant energy (NVE)

With this ensemble, no temperature, volume, or pressure scaling is done. Note that with this ensemble type, the total energy may not be conserved if cutoffs on nonbonded interactions are used (as will often be the case), or if too long an MD time step is used. In most cases, failure to conserve energy will lead to an unstable MD simulation.

Constant pressure (NPT)

For this ensemble, both temperature and pressure are held constant during the simulation. This is accomplished by also coupling the system to a pressure bath using the algorithm of Berendsen, et al. (*J. Chem. Phys.* **1984,** *81,* 3684). Seven settings become available when this ensemble type is chosen:

- Target temperature. The same as for the Constant temperature ensemble.
- Temperature relaxation time. The same as for the Constant temperature ensemble.
- Target pressure. This text box specifies the desired pressure in atmospheres. The actual pressure will fluctuate about the desired value. At each MD step the system

will be scaled such that the pressure will approach the desired value on a timescale determined by the Volume Relaxation Time parameter described below. The default pressure value is 1 atm. The acceptable range is any positive or negative real value.

- Volume relaxation time. This text box sets the volume relaxation time in picoseconds for volume scaling in a constant pressure MD simulation. The default value is 0.01 ps. The acceptable range is any value greater than zero.
- Solvent isothermal compressibility text box. Isothermal compressibility or κ (*1/V(dV/dP)*, in units of atm^{-1}) is the pressure analogue of the heat capacity and relates to the tendency of the solvent's volume to increase or decrease during pressure fluctuations in the system. The default is the value for water: 4.96 x 10^{-5} atm^{-1}. The acceptable range is any value greater than zero.
- Effective density. This text box specifies the effective density (g/cm^3) of the solute molecules. This quantity is used to compute long-range corrections to the pressure during NPT molecular-dynamics simulations. The default value is 1.0 g/cm^3. The acceptable range is any value greater than zero.
    - Volume scaling. Select one of two options for volume scaling in a MD simulation. Molecule center of mass (default). This method of volume scaling is best for small molecules and is implemented by scaling the coordinates of the center of mass for each molecular species relative to the center of the simulation box.
    - Atom based. This method of volume scaling is best for large molecules and is implemented by uniformly scaling all atomic coordinates relative to the center of the simulation box.

Initialize velocities from gaussian distribution

Selecting this option initializes the velocities of the molecules from a Gaussian distribution. The Initial temperature text box becomes available when this button is selected.

Initial temperature

This text box is available for all three ensembles when Initialize velocities from gaussian distribution is selected. It sets the initial temperature for the MD simulation. The default value is 298.15 K. The acceptable range is any value greater than or equal to zero K.

Chapter 11: Hybrid Monte Carlo Simulations

11.1 Using the Hybrid Monte Carlo Panel

Use the Impact Hybrid Monte Carlo panel to set up and run a Hybrid Monte Carlo (HMC) simulation on the Workspace structure. HMC simulations achieve relatively efficient sampling by interleaving Monte Carlo moves with a short sequence of molecular-dynamics steps. Because HMC is used mainly as a sampling method (for example, in Liaison binding affinity calculations), the MD steps can use a somewhat larger time step than would normally be advisable. The Metropolis algorithm determines which MC moves should be accepted or rejected. This ensures that the simulation does not go far astray, even if the MD time step would normally lead to a failure of energy conservation; this is why HMC is sometimes called "bad MD but good MC."

To open the Impact Hybrid Monte Carlo panel:

- In the Maestro Applications menu, choose Hybrid MC from the Impact submenu.

In the upper part of the panel are the standard FirstDiscovery panel options for Job name, Login, and Host, as well as Source of job input and Incorporate output into project by. For a description of these options, see Section 1.2. The default job name for Impact Hybrid MC jobs is `impacttmp`.

The input for an Impact Hybrid MC job can be either the contents of the Workspace or a single entry in the Project Table.

To perform Hybrid Monte Carlo simulations on a system composed of multiple entries:

1. Include only those entries in the Workspace.
2. Select Workspace as the Source of job input.

11.2 Impact Hybrid Monte Carlo Panel Features

The Impact Hybrid Monte Carlo panel has four tabbed folders:

- Potential
- Constraints
- MD Parameters
- HybridMC

The Potential and Constraints folders are described in Chapter 9. Additional features of the Constraints folder appear in Section 10.3 on page 166. Features of the MD Parameters folder relevant to both Impact Dynamics and Impact HMC are discussed in Section 10.4 on page 167. The HybridMC folder is discussed in the next section.

11.3 The Hybrid MC Folder

The HybridMC folder defines the basic settings of the HMC task. The selection options are:

Number of HMC cycles

This text box sets the number of HMC cycles for the simulation. The default value is 100. The acceptable range is any number greater than zero.

MD steps per HMC cycle

This text box sets the number of MD steps per HMC cycle for the simulation. The default value is 4. The acceptable range is any value greater than zero. Liaison calculations use 5 MD steps per HMC cycle.

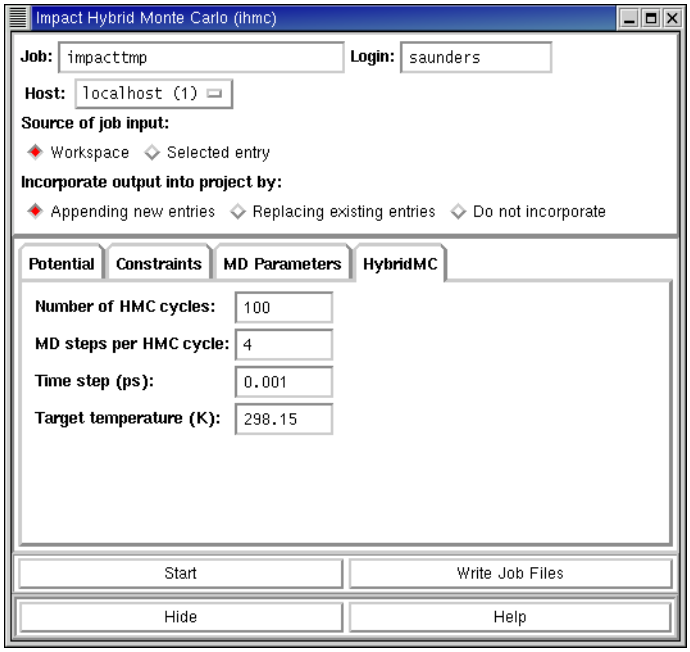


Figure 11.1. The HybridMC ***folder of the*** Impact Hybrid Monte Carlo ***panel.***

Time step

This text box sets the MD global time step (in picoseconds) for the simulation. The default value is 0.001 ps. The acceptable range is any value greater than 0 ps. Because energy conservation is less important in HMC simulations, a time step of 0.002 ps or greater may be suitable.

Target temperature

This text box sets the target temperature (in Kelvin) for the HMC simulation. The default initial temperature is 298.15 K. The default target temperature is also 298.15 K. The acceptable range is any value greater than 0 K.

Chapter 12: Soak—Add Explicit Water Solvent

12.1 Using the Soak Panel

Soak surrounds the molecule or molecules currently in the Workspace with a box of solvent molecules. You specify the box size and the requested solvent density. Then Impact adds the solvent molecules, removes any that are too close to the solute, and writes out a Maestro-format file, *jobname*_out.mae, for the soaked system. The soaked output structure will automatically be loaded into the Project Table, if you submitted the job from a Maestro project and monitor the job while or after it completes. Alternatively, you can import this file into Maestro manually or use it for another purpose.

To open the Soak panel:

- In the Maestro Applications menu, choose Soak from the Impact submenu.

In the upper part of the panel are the standard FirstDiscovery panel options for Job name, Login, and Host, as well as Source of job input and Incorporate output into project by. For a description of these options, see Section 1.2. The default job name for Soak jobs is impacttmp.

The input structure for a Soak job can be either the contents of the Workspace or a single entry in the Project Table.

To run Soak on a system composed of multiple entries:

1. Include only those entries in the Workspace.
2. Select Workspace as the Source of job input.

To model active sites or water shells around proteins with explicit solvent:

1. Run Soak.
2. Run a short minimization and constant temperature MD equilibration at room temperature.

By default, Soak places 216 water molecules in the smallest permitted solvent box (at least 18.62 Å in each dimension.) The resulting box is not equilibrated due to edge effects. A short minimization is usually sufficient to obtain a fully equilibrated solvated system. It is recommended that constant temperature molecular dynamics, described in the "Impact Dynamics Panel" and "Impact Dynamics Folder" help topics, be used for this and any other explicit solvent systems.

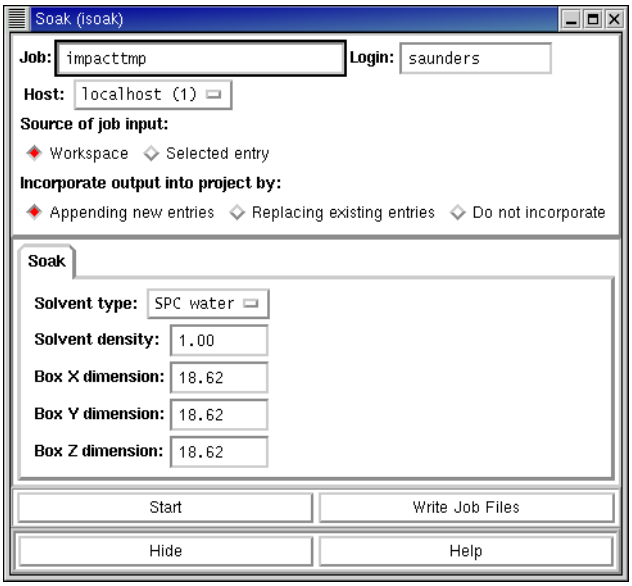


Figure 12.1. The* Soak *panel.

12.2 Soak Panel Features

The Soak panel has one tabbed folder with the following selection options:

Solvent type

At present the only solvent type that can be added using Soak is SPC (Simple Point Charge) water.

Solvent density

This text box specifies the density of solvent to be placed around the solute (units = g/cm^3). The default is 1.00.

Box X dimension
Box Y dimension
Box Z dimension

Use these three text boxes to specify the required size of the solvent box in Å. The minimum size for any dimension that Maestro will use is 18.62 Å. These values also update the size of the simulation box when periodic boundary conditions are applied. (See the discussion of periodic boundary conditions in Section 9.4.5 on page 158.)

Chapter 13: Getting Help

For help installing and setting up licenses for Schrödinger software, see the *Schrödinger Product Installation Guide*.

The Maestro help facility consists of Auto-Help, Balloon help (tooltips), and online help. To get help, follow the steps below.

- Check the Auto-Help window located below the title bar of the main window. If help is available for the task you are performing, it is automatically displayed there.
- If your question concerns an interface element, e.g., a button or option menu, there may be Balloon help for the item. Move the mouse pointer over the element. If there is Balloon help for the element, it appears within a few seconds.
- If you do not find the help you need using the steps above, click the Help button in the panel for whose settings you are seeking help. The Help panel is opened and a relevant help topic is displayed.
- For help with a concept or action not associated with a panel, open the Help panel from the Help menu on the main menu bar or by using the key combination ALT+H.

If you do not find the information you need in the Maestro help system, check the following sources:

- The *Maestro User Manual* for questions about Maestro
- The *FirstDiscovery Technical Notes* for information about technical or scientific issues
- The *Maestro Release Notes*
- The *FirstDiscovery Release Notes*
- The Frequently Asked Questions page, located at http://www.schrodinger.com/Support/faqs.html

The manuals and the release notes are available in PDF format from the Schrödinger web site at http://www.schrodinger.com/Support/pdf.html. Information on additions and corrections to the manuals is also available from this web page.

If you have questions that have not been answered from the above sources, contact Schrödinger using the information below.

Schrödinger
E-mail: help@schrodinger.com
USPS: 1500 SW First Ave. Suite 1180, Portland, OR 97201
Phone: (503) 299-1150
Fax: (503) 299-4532
WWW: http://www.schrodinger.com
FTP: ftp://ftp.schrodinger.com

Generally, e-mail correspondence is best because you can send machine output, if necessary. When sending e-mail messages, please include the following information, most of which can be obtained by entering `$SCHRODINGER/machid` at a command prompt:

- All relevant user input and machine output
- FirstDiscovery purchaser (company, research institution, or individual)
- Primary FirstDiscovery user
- Computer platform type
- Operating system with version number
- FirstDiscovery version number
- Maestro version number
- mmshare version number

Index

Numerics

3D ligand structures 59
5– and 6–membered rings 65

A

active site
    metal in 72
    with metal atoms 69
acyclic torsion bonds 83
Add hydrogens toolbar button 59
adding hydrogens 59
algorithm
    conjugate gradient 163
    energy minimization 163
    integration 167
    SHAKE/RATTLE 160
    steepest descent 164
alpha parameter, in Ewald method 159
amide bond rotation 83
annealing 68
applyhtreat 59
`applyhtreat`, brief description 60
atom constraints 160
    Liaison 160
    not used in Glide 160
    QSite 142, 160
Atom Selection dialog box (ASD). 21, 80, 131
atoms
    buffered 131, 162
    frozen 131, 161
    harmonically restrained 162
    selecting 21–22
Auto-Help 26, 179

B

backbone cuts 150
backbone picking, residue selection by 150
Balloon help 26, 179
basis set 152
batch processing, of ligands 110
binding affinity 68
binding energy model
    GlideScore 114
    in Liaison 113
    LRM 113
bond constraints 160
bonds, rotatable 65
boundary conditions, periodic 158
bounding box 76
box
    bounding 76
    enclosing 75, 76
buffer force, setting 162
buffered atoms 131, 162
buffered-atom constraints 142, 160
    QSite 142, 160
Build panel 19
building structures 18–20
button menu 8

C

Cartesian coordinate constraints 142
center, of ligand 76
centroid, ligand 76
ChemScore 68, 86
close contacts 69, 77
clustering 90
coarse grid 88
Collect MD Statistics 169
command line, running jobs from 31
Command Script Editor panel 23
command scripts—*see* scripts
command-line application `protprep` 55
commands
    impact 27
    job control 34
Commands, Maestro text box 40
conformational search in Glide 65
conformations
    core 65
    generation in Glide 83
    ring 84
conjugate gradient 163
    minimization, number of steps 89
constrain all bonds 160
constraints
    buffered-atom 131, 142, 160, 162
    Cartesian 142
    frozen-atom 131, 142, 160, 161
    Glide 65, 68

Impact 130, 160
Impact bond 160
continuum solvation 156
methods 159
QSite 141, 148
SGB model 113
conventions, document 2
convergence criteria 164
core conformations 65
core, ligand 65
Coulomb energy 68
Coulomb-van der Waals, score 90
current energy 89
current working directory 6
cutoff distance, residue-based 157
cuts 150
backbone 150
QSite 147
side-chain 150

D

database screening 68
`desalter` (LigPrep utility) 60
diameter, ligand 66
dielectric constant 155
distance-dependent 155
setting value 156
directory
current working 6, 23
FirstDiscovery 27
Impact 27
Maestro 27
output 23
structure, Liaison 27
utilities 35
distance-dependent dielectric 89
distributed processing 33, 65
Glide 32, 82, 110
Liaison 121, 136
docking
extra-precision 69
flexible 66, 83
flexible receptor 77
rigid 66, 84, 93
rigid receptor 69
SP mode 69
XP mode 69
docking mode 70, 71, 72
SP 69
XP 69
duplicate binding sites 43
duplicate chains 42
duplicate poses 90
Dynamics folder 170
dynamics simulations 165

E

electrostatic grids 68
electrostatic treatment 154
constant dielectric 155
distance-dependant dielectric 155
Emodel 68, 87
enclosing box 75, 76, 96
end of ligand file 83
energy
current 89
internal 87
internal strain 68
nonbonded interactions 87
energy change criterion 164
energy grid score 68
energy minimization
Impact 162
steric clashes in 164
enrichment factor 68
entries, Project Table 12
including, excluding, and fixing 16
merging 139
selecting 15–16
sorting 14
environment variables
`DISPLAY` 5–6
`SCHRODINGER` 5–6
ePlayer 13, 14
Equations of Motion (EOM) 167
Ewald long-range correction 159
Ewald summation 140
excluded entries 16
explicit solvent 165
extra-precision
(XP) mode 69
Glide 65

F

Fast Multipole Method (FMM). 140, 156, 157, 163
  hierarchy level setting 158
  and RRESPA 167
file I/O directory 23
file names, in FirstDiscovery 28
file, parameter 154
filters 15–16
  hierarchical 86
final scoring 89
five– and six–membered rings 83
fixed entries 16
flags, Jaguar 148
flexible docking 66, 68, 83
flexible ligands 66
flexible receptor docking 77
force field 28, 154
  QSite 140
format conversion, to Maestro 59
format, Mol2 81
fragments, building structures with 18
free ligands 151
free ligands selection 150
free ligands selection mode 151
frozen atoms 131, 161
  in Impact constraints 161
frozen-atom constraints 142, 160
  QSite 142, 160

G

generate conformations 83
Glide
  Active Site Residues panel 79
  conformational search 65
  Constraints folder 68, 94
  described 65
  distributed processing 32, 82, 110
  Dock Displayed Ligand 80
  Glide panel in Maestro 70
  Ligand folder 80
  Ligand to Define Grid panel 78
  Output folder 90
  panel 65
  Pose Viewer 103
  PoseWrite panel 107
  Reference Ligand panel 85
  Scoring folder 86
  Settings folder 70
  Similarity folder 98
  Site folder 74
Glide constraints 65, 69
Glide docking
  mode 72
  standard precision 72
  subjobs 110
Glide XP mode 65, 69
Glide/Prime induced fit 77
glide_rescore utility 109
glide_sort utility 93, 108
GlideScore 68, 86, 89
  binding energy model 114
  in Liaison 114
gradient criterion 164
greedy scoring 67
green crosses 152
grids
  coarse 88
  electrostatic 68
  receptor 67, 70
  smoothed 68
  van der Waals 68
grow bond 18

H

Help button 179
help option (-h) 35
Help panel 26, 179
Hide button 5
hierarchical filters 86
hydrogen treatment 59
hydrogen-bond score 90
hydrogen-bonding interaction 95

I

identical chains 43
identical sequences 42
Impact atom constraints 160
Impact bond constraints 160
impact command 27, 29
Impact constraints 160
  buffered-atom 131, 162
  frozen-atom 131
Impact Dynamics panel 165
Impact energy minimization 162
implicit solvent models 156, 159

import structure 39
impref utility 37, 58
included entries 16
induced fit 77
Induced Fit protocol 69
initial step size, of minimization cycle 164
in-place scoring 84, 93
integration algorithm
  in MD 167
  RRESPA 167
  Verlet 167
interface behavior, of Maestro 5
internal energy 87
internal strain energy 68
ionization state expander 63
ionizer utility 63
ions, QM region 151

J

Jaguar 139
  flag settings 148
  keywords 148
  parallel processing 148
job control commands 34
jobname.trj 169
jobs, running 31, 34
jobs, running in Maestro 24–25

K

keywords, Jaguar 148
killing Liaison jobs 137
K-space vectors, maximum length 159

L

levels, number of in FMM 158
Liaison
  Analysis 118
  Analysis folder 132
  Constraints folder 130
  described 113
  directory structure 27, 115
  distributed processing 121, 136
  fitting simulation results 118
  Impact atom constraints 160
  job files created 116
  killing jobs 137
  methodology 113
  multiple processors 121, 136
  panel in Maestro 119
  Parameters folder 125
  predicting binding affinities 119
  running from the shell 136
  running remotely 114
  running through Maestro 114
  Settings folder 120
  System folder 122
Liaison (Impact) constraints 130
license, LigPrep 60
ligand-receptor interaction (Glide constraint) 68
ligands
  center 76
  centroid 76
  diameter 66
  flexible 66
  free 150, 151
  nonpolar atoms 84
  poses 65
  protein preparation 49
  reference 85
  residue selection 150
  rigid 66
  structure requirements 59
`ligparse`, brief description 61
LigPrep 59–63
Linear Interaction Approximation (LIA) ... 113
Linear Response Method (LRM) 113
Linear Synchronous Transit (LST) 146
list, neighbor 157
log file, saving (Maestro) 26
long range forces, in TN optimizations 164
LRM binding energy model 113

M

Maestro
  Commands text box 40
  help 179
  interface behavior 5
  main window 6, 7
  menus 7–8
  quitting 26
  running jobs from 24–25
  scratch projects 12
  starting 6
  undoing operations 24
  working directory 27

main window 7
maximum step size of minimization cycle. 164
MD Parameters folder 167
MD simulations, frequency of printing information 169
menu button 8
Merge entries 139
metal in active site 69, 72
metal ions in receptor 95
metal-ligand interaction 95
metal-ligand score 90
minimization
  conjugate gradient 163
  convergence criteria 164
  cycles, maximum 163
  Truncated Newton 163, 164
Minimization folder, QSite 143
MM potential energy 140
MM region 142, 160
Mol2 format 81
molecular dynamics (MD) 165
molecular mechanics
  electrostatic treatment 154
  parameter file 154
Monitor panel 25
mouse functions 5
  Project Table panel 16–17
  Workspace 11
multimer 42, 43
multimeric protein structure 38
multiple processors 33, 148
  Glide 110
  Liaison 121, 136

N

neighbor-list frequency, updating 157
`neutralizer`, brief description 61
nitrogen centers, trigonal 83
nonbonded interaction energy 87
nonpolar atoms
  ligand 84
  receptor 77

O

online help 26, 179
OPLS1999 ..... 28, 140
OPLS2000 28
OPLS2001 28
OPLS-AA 67, 68, 154
overview of protein preparation 38

P

para_glide utility 32, 65, 82, 110
parallel processing 33
  and continuum solvation 148
  Jaguar 148
  QSite 141
  See also distributed processing
parameter file, molecular mechanics 154
periodic boundary conditions 140, 156, 158, 165
pick states 8, 21
picking controls 80
Poisson Boltzmann Solver (PBF) 159
Pose Viewer, *See* Glide
pose, ligand 65
poses per ligand, maximum number
  initial 88
  refined 88
potential energy
  MM 140
  QM 140
Potential folder 154
pprep utility 37, 57
Preferences panel 23, 24
Prime 69
processors, multiple 148
product installation 179
project entries 12
Project Facility, introduction 12
Project Table panel 13
  menus 15
  mouse functions 16–17
  shortcut keys 17
projects 12
protein
  multimeric 42
  preparation for pprep, overview 38
protein complex structure
  importing 38, 39
  truncating 42
protein preparation
  incorporate results as 53
  overview 38
protein preparation facility 37
protein structure, multimer 38

protprep command-line application 55
pyramidalization 83

Q

QM options 148
QM potential energy 140
QM region 142, 160
    ions 151
QM/MM 139
QM/MM boundary 150
QSite
    atom constraints 142, 160
    continuum solvation 141, 148
    cuts 147
    description 140
    force field 140
    Minimization folder 143
    parallel processing 141
    QM Region Ions panel 151
    QM Residues/Ligands panel 149
    QM Settings folder 147
    QSite panel 140
    single-point energy 145
    unavailable methods 140
Quadratic Synchronous Transit (QST) 146
quantum mechanical calculations 139
quasi-Newton method 146
quitting Maestro 26

R

range of ligand structures 83
range value 83
raw (unsorted) pose file 92
receptor constraint sites 69
receptor grids 67, 70
Record Trajectory 169
reference ligand 85
region boundaries, QSite 147
reject pose unless 89
residue selection for QM region 150
residue-based cutoff distance 157
Reversible REference System Propagator Algorithm— *see* RRESPA
rigid docking 66, 84, 93
    ligands 66
    rigid receptor 69
ring flips, allowing 84
ring_conf, brief description 61
rings 65
rms distance 85
rotamer groups, ligand 65, 66
rotatable bonds 65
rough score 88
    screening 89
RRESPA integration algorithm 167

S

Scale radius if 84
Scaling of vdW radii 77
Schrödinger contact information 180
$SCHRODINGER/utilities directory 93
score
    Coulomb-van der Waals 90
    hydrogen-bond 90
    metal-ligand 90
    rough 88
scoring
    final 89
    in-place 93
scoring function 68, 87, 88
scoring in-place 84
scratch entries 12
scratch projects 12
screening, rough score 89
scripts
    command-line 35
    Maestro 22
    utility 35
sdconvert, LigPrep use of 60, 62
Sequence Viewer 42
sequences, identical 42
SET FFIELD 28
SGB— *see* Surface Generalized Born (SGB)
SHAKE/RATTLE algorithm 160
shortcut keys
    main window 11–12
    Project Table panel 17
side-chain cuts 150
side-chain picking, residue selection by 150
similarity 65
similarity scoring 98
    standard 98
    weighted 98
single-point energy calculation 145
smoothed grids 68

solvation method 159
solvation, continuum 141, 148, 156
solvent molecules 150
solvent, implicit 156
standard TS optimization 146
standard-precision (SP) docking mode 69
standard-precision Glide docking 72
steepest descent 164
`stereoizer`, brief description 61
steric clashes, in energy minimization 164
Stop Overall Motion 169
strain energy 68
structure format conversion 59
structures
  building 18–20
  displaying in sequence 13
subjobs, Glide docking 110
subset test, Glide 66
summation method, Ewald 159
Surface Generalized Bohr (SGB) 140
Surface Generalized Born (SGB) 113, 159
symmetry-equivalent atoms 95

T

`tautomerizer`, brief description 61
Technical Notes 68
technical support 26, 179
toolbar 39
  Build panel 20
  main window 8–11
  Project Table panel 14
trajectory analysis 169
trajectory file 169
  analysis 28
transition-state optimization
  LST 146
  QST 146
  Standard 146
trigonal nitrogen centers 83
Truncated Newton (TN) algorithm 163, 164
Truncated Newton minimization 143
truncation
  nonbonded interactions 141, 156
  protein complex structure 42

U

undoing Maestro operations 24
Use Smoothing 158
utilities
  command-line 35
  glide_rescore 109
  glide_sort 93, 108
  help option (-h) 35
  impref 37, 58
  ionizer 63
  para_glide 32, 65, 82, 110
  pprep 37, 57
  scripts 35

V

van der Waals grids 68
van der Waals radii 69, 77, 84
Verlet integration algorithm 167

W

water entry, preparing 39
waters
  crystallographic 39
  deleting 39
  in protein complex structure 39
  structural 39
working directory, Maestro 27
Workspace
  description 6
  including, excluding, and fixing entries 16
  mouse functions 11
  scratch entries 12
Write Template 28

X

XP Glide 65
XP mode 69

120 West 45th Street
32nd Floor
New York, NY 10036

1500 SW First Avenue
Suite 1180
Portland, OR 97201

3655 Nobel Drive
Suite 430
San Diego, CA 92122

Dynamostraβe 13
68165 Mannheim
Germany

SCHRÖDINGER